An Observational Study of Galaxy Clustering

Paul Charles Hewett

Doctor of Philosophy

University of Edinburgh

1982

To Caroline

This thesis has been composed by myself and consists entirely of my own work except where specifically indicated in the text.

November 1982

Contents

Abstract

An observational study of the clustering of faint galaxies in the magnitude range $19.5 < m_J < 22.0$ is undertaken. High quality plate material from the United Kingdom Schmidt Telescope, and automated machine measures from the COSMOS facility at the Royal Observatory Edinburgh constitute the primary data. Particular attention is paid to quantifying the amplitude and nature of random, and systematic errors associated with the plate material and the measuring process. The details of calibration procedures and properties of the COSMOS scanning spot are found to be particularly important for a full understanding of the measurement process. New information concerning the success of automated image classification techniques is obtained, through the use of very deep plate material from the Anglo-Australian Telescope to "calibrate" the wide field plate material used in the investigation. Automated image classification using COSMOS parameters is found to be considerably less successful than previously claimed. Consideration of the methods used to estimate the covariance function from galaxy samples shows that some discrepancies between earlier work on the Zwicky and Lick catalogues, and more recent, deep machine measured samples are due to different estimation procedures, and not the data themselves.

Data from five high galactic latitude fields in the south, subtending a total area of more than 100 square degrees, allows precise limits to be placed on the amplitude of spurious galaxy clustering due to small scale extinction fluctuations within the galaxy. Upper limits to the amplitude of extinction fluctuations are derived from the form of the stellar covariance function and the cross-correlation between the stellar, and deep galaxy samples. The observations are compared to a number of extinction models, and it is concluded that mean values of absorbtion at

the poles of $A_B \leq 0.1$ magnitudes are favoured.

Analysis of the galaxy distribution in the five fields shows that the heirarchical model for the distribution of galaxies that arises naturally from the gravitational instability theory of galaxy formation, matches the observations well. However, the small angular extents of the five fields means that constraints on the amplitude and nature of very large scale features are weak. Clustering models involving filamentary and sheet like structures, consistent with the adiabatic theory of galaxy formation, can also match the data. In both cases structure on scales of at least $10h^{-1}$Mpc is necessary to explain the observations, but because of the low signal to noise of the clustering in the samples, discrimination between the models is not possible.

Acknowledgements

I would particularly like to thank the following:

David Emerson for continuous help and encouragement throughout the project

Sue Tritton for her invaluable help in the plate library and assistance with image classifications

Steve Heathcote for the use of his graphics package on the ROE GEC 4090

The staff of the IDPU and UKSTU units at the ROE for plates and measurements

The following people who have provided help, advice and data at various stages;

Malcolm Longair, Simon Lilly, Peter Brand, Adrian Melott, Keith Tritton, Bob Stobie, Bernie McNally, Dennis Kelly, John Cooke, Nigel Sharp, John Peacock, Arthur Trew, Gerry Gilmore, Neill Reid, Susan Hooper, Len Lawrence, Harvey MacGillivray, Morag Brown, Dick Fong, Tom Shanks, Richard Ellis, Jon Godwin, John Dawe, Nigel Metcalfe, Richard Prestage, Susan-Jane Clement, Harold Corwin, Steven Beard, Clive Davenhall, Mike Hawkins, Alison Campbell, Liz Sim, Ralf Martin, Dorothy Skedd and Russell Cannon.

1 Introduction

Progress over the past two decades in development of observational techniques, and accumulation of data necessary to constrain theories of galaxy formation has been rapid. Discoveries such as the detection of the three-degree microwave background (Penzias and Wilson 1965), the strong redshift evolution of the radio source (Peacock and Gull 1981), and quasar number counts (Wall 1980), have enabled considerable confidence to be placed in the standard big-bang evolutionary cosmology (Weinberg 1972, Gunn, Longair and Rees 1978). The framework to which theories of galaxy formation must address themselves has thus become much better defined. Recently the connection between advances in fundamental particle physics and conditions in the very early universe have suggested that a solution to the most intractable of problems - the origin and amplification of seed fluctuations - may be in sight (e.g. Ellis J. 1982, Guth 1981, Gott 1982). Detailed theoretical treatment of the conditions in the very early universe however await considerable advances in theoretical and experimental aspects of particle physics.

Considerable progress has also been made in gathering observations relating to galaxy formation. Much effort has been concentrated in mapping the amplitude, and size spectrum of fluctuations in the radiation and matter components of the visible universe. In parallel, numerous theoretical attempts to relate these observations to the initial conditions of galaxy formation or subsequent interactions between matter and radiation have been made (e.g. Peebles 1981, Press and Lightman 1978, Doroshkevich et al. 1978). Quantification of the observed matter and radiation fluctuations, coupled with an understanding of temporal evolution will constrain the conditions pertaining in the early universe that gave rise to galaxy formation.

Success in determining the nature of fluctuations in the observed radiation field has been confined to the microwave background radiation (Fabbri et al. 1980, Boughn et al. 1981). Searches for fluctuations at other wavelengths which may also probe far back in time have yielded only upper limits (e.g. Davis 1981, Rees 1981), and the precision required to discriminate between theoretical models of galaxy formation is at the limit of current technology at all wavelengths. The microwave fluctuation results themselves are still uncertain (Wilkerson 1982).

Similar problems arise in mapping the fluctuations of the largest structures of luminous matter - galaxies and clusters of galaxies. In this case it is necessary to assume that the luminous matter is a good tracer of the underlying mass distribution. Vigorous debate continues over the relation between luminous and possible "dark" matter (Ostriker et al. 1974, Davis et al. 1980, Faber and Gallagher 1979), and practically all work on the clustering of galaxies assumes a high degree of correlation; if this is not the case then conclusions drawn from these studies may be misleading. In this study it will be assumed that observations of the galaxy distribution do reflect the distribution of the majority of the mass in the universe - at least on scales larger than a single galaxy.

Even with this assumption, the mapping of the observed mass distribution well enough to discriminate between galaxy formation theories is beset with problems. For technical and observational reasons current statistical quantification of the galaxy distribution is confined to a small range of linear scales where the clustering of galaxies is intrinsically strong and therefore easy to detect. The most precise data is also confined to relatively nearby regions of the universe (Davis et

al. 1982, Yahil et al. 1980), and it is not clear what fraction of the universe needs to be studied to provide a "fair sample". If the study of galaxy distributions is to prove useful in constraining galaxy formation theories, great improvements in the quality and quantity of the data are required. In particular the quantification of galaxy clustering must be extended to larger scales and made far more precise; currently several theories can match the observations because of the limited range in linear scale and low precision of the observations.

The available galaxy catalogues: the wide angle Zwicky catalogue (5000 galaxies, effective depth $50h^{-1}$ Mpc (Zwicky et al. 1961-1968)) and Lick catalogue (800,000 galaxies, effective depth $220h^{-1}$ Mpc (Shane and Wirtanen 1967)) together with the narrow angle Jagellonian catalogue (12,000 galaxies, effective depth $380h^{-1}$Mpc (Rudnicki et al. 1973)) were laboriously constructed from eye scans of - by modern standards - poor plate material. Numerous selection effects, substantial random errors, together with large scale non-uniformities are present in the catalogues. (Throughout h denotes the value of Hubble's constant "H" in units of 100 km $s^{-1}Mpc^{-1}$.) These effects may be crudely removed (Seldner et al. 1977) but for high precision large scale clustering quantification little information can be reliably extracted. The initial aim of this project was to achieve the desired improvements in data quality by obtaining a well defined sample of galaxies, with high precision parameters, from a volume larger than any yet surveyed.

For a new sample to be able to extend the scale, and increase the precision of the clustering estimates, order of magnitude improvements in the determination of galaxy parameters, - e.g. positions, shapes and magnitudes - and the uniformity of detection criteria within the catalogue are necessary. Even with the vast increase in detector

efficiency in recent years such a sample has to be confined to two-dimensional (2D) catalogues; sufficient redshifts to provide a three-dimensional (3D) catalogue are not yet feasible, although considerable advances in objective prism redshift techniques (Cooke et al. 1981) may make this possible in future. A 2D sample of ~500,000 galaxies with an effective depth approaching $1000h^{-1}$Mpc surveying a volume of five million cubic megaparsecs would enable quantification of galaxy clustering to scales of $30h^{-1}$Mpc, and provide a severe test of the scaling behaviour and small scale clustering results derived from earlier surveys.

Such a project requires three major components; a large collection of deep high quality, sky-limited, wide angle Schmidt plates, a fast microdensitometer to perform plate digitisation and preprocessing of the vast quantity of data, (10^9bits per plate) and finally the computer time and software techniques for construction and subsequent analysis of the galaxy catalogue. In principle the facilities for such an investigation are available; a large library of original United Kingdom Schmidt Telescope Unit (UKSTU) sky-limited J plates with a limiting magnitude of m_J=23 (UKSTU handbook 1981), the COSMOS measuring machine, capable of digitising the central ~25 square degrees of an UKSTU plate, with subsequent image detection and parameterisation performed within 24 hours (Stobie et al. 1979). Preliminary work on software techniques has also taken place - e.g. image classification (Carter and Godwin 1979) and computer modelling of galaxy fields (Soneira and Peebles 1978).

Achieving the requisite improvements in accuracy however places strict performance limits on the constituent components of the project. The advantages of a 2D sample are the large number of galaxies, and size of region that can be rapidly surveyed due to the enormous information

gathering properties of a Schmidt-telescope/photographic-plate combination. The main disadvantage of working with a 2D catalogue is the washing out of spatial galaxy clustering by the overlapping of many clusters at different distances seen in projection. As a result large amplitude spatial features appear as only weak density enhancements in the 2D projections. To successfully detect the boundary of a non-linear (i.e. density contrast unity at its edge), spatial feature (due to superclustering say), with overlapping galaxies and clusters along the line of sight providing 90% of the total galaxy counts, 10% number density enhancements must be detectable. This allows crude limits to be placed on the maximum amplitude of the systematic effects in the final galaxy catalogue: fluctuations present at scales of $10h^{-1}$Mpc subtend angles of order a degree ($10h^{-1}$ Mpc subtends one degree at Z=0.2), integral number counts of galaxies at m_J=21 (magnitude of galaxy at the knee of the luminosity function at Z=0.2) show a magnitude dependence approximating Log(number)$\propto$ 0.45*M (Ellis R. 1982). Variations in the detection criteria, and magnitude scale for galaxies must be confined to $\leq$0.1 magnitudes on scales of order a degree if spurious number-density fluctuations are to be kept below the level expected due to intrinsic galaxy clustering. Random errors must also be kept to a minimum; large random errors in galaxy magnitudes decrease the signal to noise, smearing the already broad Number(galaxies) vs. Redshift selection function, making density fluctuations that much harder to detect.

These severe performance criteria apply to the plate material/measuring machine/ software-reduction combination; the limitations for each component are correspondingly stricter. Although each of the available components is manifestly better than any previously employed it is not clear that they meet these criteria.

Subsequent to the compilation of a preliminary object list from the raw data, the same fluctuation detection limitations apply to the software techniques; image classification must have a high probability of success (to keep the random contamination from stars down), and show no variation in success rates with plate position or other factors. Differential number counts of stars and galaxies are approximately equal at m_J=20.5 (e.g. Koo and Kron 1982), and a variation in classification success probability of ~10% over scales of a degree will again make quantification of low amplitude features impossible.

Unavoidable extrinsic clustering effects must also be removed from a 2D galaxy catalogue. A major difficulty is correcting for the apparent galaxy clustering due to the variable optical depth of dust along different lines of sight through the galactic plane. Debate over the importance of variable extinction on current catalogues continues (e.g. Holmberg 1974, Burstein and Heiles 1978, Strong and Lebrun 1982). Assuming a simple Spitzer standard cloud model for the interstellar medium (Spitzer 1978), fluctuations of order 0.2m in detection limit can be expected on scales of order a degree - comparable to the amplitude of the intrinsic density fluctuations at this scale in the galaxy catalogue.

Once a galaxy catalogue is obtained, clustering statistics must be capable of identifying the key features in the distribution. Argument persists over whether commonly used clustering statistics (e.g. the correlation functions) are capable of identifying filamentary structures in the galaxy distribution, or even between relatively simple models for the galaxy distribution (Shanks 1979, Peebles 1979) - filamentary or pancake like structures are one of the features of the adiabatic galaxy formation scenario (e.g. Zeldovich 1978). As simple statistics reduce

the positional information of every galaxy and its relation to the rest of the catalogue to a few numbers, it is not always obvious the correct information is being distilled from the catalogues.

The bulk of this thesis is concerned with obtaining the precise performance limitations of the components involved in construction of a 2D galaxy catalogue as outlined above, and developing techniques (that were not available) to perform several of the offline analysis tasks. Chapter 2 reviews the current state of observational galaxy clustering, attempting to clarify the somewhat confused literature concerning the nature of large scale structures in the observed galaxy distribution. Chapters 3 and 4 deal with the quantification of the performance limitations of the two major components of the investigation - UKSTU wide field photographic plates and the COSMOS machine. Chapter 5 describes the results obtained in the field of image classification using COSMOS data together with the finally adopted classification procedures. Chapter 6 deals with the estimation of the prime galaxy clustering technique in use today - the covariance function - showing how the effects of bounded samples exhibiting considerable degrees of inhomogeneity may be succesfully analysed. Chapter 7 describes the procedures finally adopted for the measurement and reduction of plate material accepted for the clustering investigation. A method is developed in Chapter 8 for directly examining the effects of variable galactic extinction in a galaxy catalogue by applying clustering techniques to the usually neglected companion stellar samples. The observations are presented in Chapter 9 together with results from simulations of clustering schemes and their appearance on UKSTU plates.

2 Observations of Galaxy Clustering

2.1 Introduction

Quantitative studies of galaxy clustering were not possible until large catalogues of galaxies and galaxy clusters became available (Abell 1958, Zwicky et al. 1961-1968, Shane and Wirtanen 1967, Rudnicki et al. 1970). Prior to the compilation of these catalogues only counts of galaxies by Hubble and Shapley (e.g. Hubble 1934, Shapley 1940) provided data on "field" galaxy clustering, rather than the most extreme of galaxy clusters. The lack of suitable catalogues resulted in early studies of galaxy clustering being confined to large, obvious clumpings of galaxies - rich clusters, which themselves were not catalogued in detail until the late 1950s (Abell 1958). Observations of rich clusters have revealed much concerning the dynamics and evolutionary history of these systems, of which the "missing mass" problem (e.g. Rood 1981) has the most general consequences for cosmology (e.g. Press and Davis 1982). Properties of rich clusters have been reviewed by Bahcall (1977) and Rood (1981), but the extensive work on these systems is not directly applicable to many aspects of the general clustering of galaxies; the short crossing-times and high galaxy densities characteristic of these systems mean enviromental influences are significant within rich clusters (e.g. Chamaraux et al. 1980, Himmes and Biermann 1980), relating observed properties of rich clusters to conditions pertaining at their formation epoch may prove difficult or impossible.

The preoccupation with large rich systems was at the expense of smaller, less obvious, but far more common galaxy clusters. Work on these systems has been hampered by the difficulty of identifying genuinely associated galaxies, and deriving an accurate treatment of the selection procedures

involved (e.g. Gott and Turner 1976, Turner and Gott 1976, cf. Appleton and Davis 1982). Only recently has it been possible to obtain sufficient redshift information that large samples of small groups of galaxies can be selected and statistical studies undertaken (Tully 1982, Huchra and Geller 1982).

Rich cluster studies and recent analysis of small-group properties have detailed information on all aspects of particular systems (e.g. Tully 1980), or used cluster properties in an attempt to answer specific questions - e.g. the location of "dark matter" (Press and Davis 1982). This detailed study of small numbers of systems is complementary to a more general statistical approach where the information utilised is less complete but the resulting quantification relates to the galaxy distribution in general.

Prior to the mid-1960s quantitative data on the nature of galaxy clustering and statistical quantification of cluster properties was essentially nonexistent, exceptions being the initial study of the distribution of Abell clusters (Abell 1961), and the preliminary analysis of part of the Lick catalogue by Neyman et al. (1953). The situation at the end of the 1960s is well reviewed by de Vaucouleurs (1971): the existence of galaxy clustering on all scales up to $10h^{-1}$ Mpc was recognised, with indications that larger scale structures existed in the Abell and Lick catalogues. Some workers believed that variable galactic extinction was a significant factor in modifying the observed galaxy distribution at scales greater than $5\text{-}10h^{-1}$ Mpc - an idea that persists today (Shanks et al. 1980; henceforth SFEM, Tyson and Jarvis 1980).

The early 1970s signalled the start of a programme to apply statistical techniques to describe the observed galaxy distribution (Totsuji and Kihara 1969, Peebles 1973). Two factors were crucial to this development: the availability of large catalogues of galaxies and galaxy clusters, and access to high speed computers. The most coherent approach to clustering analyses was due to Peebles and coworkers at Princeton, who developed the use of correlation techniques to describe the observed distribution of galaxies. The theory is comprehensively described by Peebles (1973, 1980), and for an excellent general review of the correlation approach see Fall (1979).

Most work has concentrated on the calculation of the two-point correlation - covariance - function, and its fourier transform, the power spectrum. To fully describe the clustering properties of a galaxy sample of n points, n moments are required. The two-point function examines only the second moment of the distribution: the pairwise distribution of objects - providing an extremely limited description of the galaxy distribution. Attempts to extend the correlation function approach to higher moments have been limited by the increased computation involved and low signal to noise of the functions in the available catalogues. The detection and behavior of the three-point function is well established for the Zwicky, Lick and Jagellonian catalogues (Groth and Peebles 1977; henceforth GP77), but the four-point function (Fry and Peebles 1978, Bonometto and Sharp 1980) gives results that can only be said not to be inconsistent with the hierarchical model suggested by the behaviour of the lower order moments (Peebles 1980).

Considerable criticism of the correlation function approach has been made (Shanks 1979, Jones 1976), particularly the limited information contained in the low order moments. Application of other techniques to

galaxy catalogues are legion, but with the possible exception of the Mead's statistic (Shanks 1979), no significant new quantitative information has been derived from these statistics not already available from the correlation functions or power spectrum. The situation may change in the future with the application of new statistics related to the higher order correlation functions (e.g. White 1979, Sharp 1981). Currently the only coherent description of the observed distribution of galaxies is in terms of the low order correlation functions.

Two of the reasons for the success of the correlation function approach are the ability to easily relate the properties of galaxy samples at different limiting magnitudes, and the simplicity of the approach, which makes the results readily comprehensible. A full review of galaxy clustering in general, and correlation functions in particular is not included here as recent reviews are available (e.g. de Vaucouleurs 1971, Fall 1979, Peebles 1980). Instead a number of specific points are discussed.

2.2 Correlation Function Observations

The analyses of the Zwicky, Lick and Jagellonian catalogues (Totsuji and Kihara 1969, GP77) showed the two-point correlation function could be well approximated by a power-law over a wide range of scales. A simple model of the form

$$W(\theta) = A\,\theta^{-\delta} \qquad (2.1)$$

with index $\delta = 0.8$, holds over scales of a tenth to ten megaparsecs. The prime conclusion is that galaxy clustering exists to scales of order $10h^{-1}$ Mpc - with no strongly preferred scales evident. An excess of clusters

of a particular size would result in a feature in the covariance function at and below scales comparable to the cluster diameter. The 2D results may be transformed to give the spatial two-point correlation function (e.g. Peebles 1980) - the quantity of physical significance - which is well represented by the model

$$\xi(r) = (r/r_o)^{-\sigma} \qquad (2.2)$$

with index $\sigma=1.8$ and $r_o=5h^{-1}$ Mpc. That is, the chance of finding a galaxy at a distance $5h^{-1}$ Mpc from another galaxy is twice that expected if the galaxies were randomly distributed. Note however, that even on these relatively small scales the two-point function varies considerably with position on the sky (GP77; Figs. 1 and 4), and the power-law fit is only an approximation to the data (e.g. Peebles 1974; Fig. 1). The power-law covariance function extends to a scale of order $10h^{-1}$Mpc, where a sharp break from the power-law behaviour is observed in the Lick catalogue. Some controversy exists concerning the existence and position of this feature (e.g. SFEM).

The two-point function gives no information on any angular dependence (i.e. shape) of galaxy clustering; $\xi(r)$ is a function of separation only. To extend the description of clustering, the three-point correlation function was calculated for the Zwicky, Lick and Jagellonian catalogues (GP77). The three-point function can be represented as a simple function of the observed two-point function (Peebles 1980; section 57), and this result has been taken to indicate there is no shape dependence in the observed galaxy distribution - e.g. galaxies do not preferentially lie in filaments. The four-point function (Fry and Peebles 1978) has been calculated for the Lick catalogue using a limited range of shapes and scales. The results are again consistent with a model in which galaxy

clustering exhibits no shape dependence. An alternative method for deriving the four-point function under a restricted set of conditions (Bonometto and Sharp 1980, Sharp et al. 1983, preprint) is broadly consistent with Fry and Peebles (1978), but the signal to noise in both cases is very low. These results led Peebles to postulate a particularly simple clustering heirarchy model for galaxy clustering that is consistent with the observed two, three and four-point correlation functions. In this scheme galaxies are distributed in a continuous clustering heirarchy with no preferred scales; the clustering looks similar at any level within the heirarchy (Peebles 1980, Soneira and Peebles 1978). The observed correlation functions can be explained by the modification of an initial white noise power-law spectrum of density fluctuations in the early universe, with subsequent evolution dependent only on gravitational interactions. This simple form of the initial density fluctuation spectrum together with one dominant form of interaction - gravity - is found to be particularly appealing by many workers.

The correlation approach has been extended to studies of Abell clusters (Hauser and Peebles 1973, Bahcall and Soneira 1983), cross-correlation between Abell clusters and Lick galaxies (Seldner and Peebles 1977) and the shape of the distribution of galaxies around Abell clusters via the cluster-galaxy-galaxy correlation function (Fry and Peebles 1980). The distribution of Abell clusters shows clear evidence for superclustering on scales up to $100h^{-1}$Mpc - the majority of superclusters consisting of pairs or triplets of clusters. The cross-correlation between the Abell and Lick calalogues revealed a strong relation between the two classes of objects, with a well defined density run of galaxies surrounding Abell clusters extending to a radius of $15h^{-1}$Mpc. An Abell cluster/halo complex is thus already close to a "supercluster" size. Considering the

distribution of pairs of galaxies about Abell clusters, Fry and Peebles (1980) failed to find a significant tendency for the clusters to be elongated or show subclustering at a stronger level than expected from the heirarchical clustering model.

In summary, the distribution of clusters and galaxies analysed by Peebles and coworkers apparently results in an extremely simple picture of clustering; the observations may be related to an initial power-law spectrum of isothermal density fluctuations through subsequent evolution of clustering by gravitational interaction only. In particular no evidence was found for galaxies to form other than symmetric structures. For these conclusions to be supported in detail, and taken as strong evidence for a simple heirarchical clustering picture three areas must be considered; (a) limitations imposed by random and systematic errors in the source catalogues, (b) the detailed scaling between catalogues at different depths, and (c) whether sufficient linear scales have been examined for the observations to severely test the heirarchical clustering model and rule out other competing schemes.

2.3 Catalogue Completeness and Selection Effects

The compilation of the Lick, Abell, Zwicky and Jagellonian catalogues as well as others that duplicate these to some degree (e.g. Nilson 1973) were mammoth undertakings. Criticism of selection procedures and detailed examination of errors as a function of catalogue position or compilation time for example are unpopular, and little published data is available. When catalogues are used for detailed quantitative statistical work however, such investigations are probably the most important if confidence is to be placed in the results.

The catalogues already mentioned were compiled by eye from what today would be considered poor plate material, and in no case has a comprehensive evaluation of the completeness, selection functions and error rates been carried out using superior external material. Internal consistency checks are possible (e.g. Shane and Wirtanen 1967) but these are no substitute for comprehensive external verification. The importance of quantifying the effects of catalogue errors on statistical analysis is illustrated by the large fraction of the Princeton analyses devoted to examining sources of bias in catalogues (e.g. Seldner et al. 1977, GP77, Hauser and Peebles 1973), and despite the careful procedures applied to the Lick catalogue, systematic effects are detectable in the statistical analysis (GP77). The overall effects of the many correction factors applied to the Lick counts is unclear. Once the multiplicative number density corrections exceed ten percent, significant modification of the data occurs, and information is clearly lost - this point has not been addressed in detail. In another important area - image classification - little consideration appears to have been given to the classification procedures and the effects of the large variation in the ratio of stars to galaxies with galactic latitude for instance - the stellar number density at the magnitude limit of the Lick counts varies by a factor three or more from $b=40^{\circ}$ to $b=90^{\circ}$.

Fesenko (1979) and Reaves (1974) have considered possible variation in the homogeneity of the Abell catalogue; in a particularly simple analysis Fesenko (1979) shows that the variance of cluster counts in sub-areas on the same plate is approximately half the variance between sub-areas on different plates. Unfortunately Fesenko's paper is not detailed enough to ascertain the significance of the result - he has not apparently performed the analysis confined to Abell's complete cluster sample. Reaves' (1974) consideration of the frequency of occurrence of

Abell clusters in plate overlap zones - which is confined to the complete sample - also indicates some degree of inhomogeneity. Consequently the amplitude of clustering found by Hauser and Peebles (1973) for Abell clusters may be too high by up to a factor two on scales comparable to a plate diameter. However the strong clustering found by Bahcall and Soneira (1983) utilising redshift information does indicate strong clustering is present.

Large random ($\pm$0.5 magnitudes) and systematic ($\geq$0.5 magnitudes) errors are present in the Zwicky catalogue (Bothun and Schommer 1982, Huchra 1976, White and Valdes 1980). The situation is mitigated by GP77's use of a bright subsample of the Zwicky catalogue, but little data exists on the reliability of Zwicky magnitudes fainter than m_B=14.5.

To summarise, the majority of available galaxy and cluster catalogues contain large random as well as unquantified systematic effects. Magnitude errors in the Zwicky catalogue (Bothun and Schommer 1982), possible inhomogeneities in the Abell catalogue (Fesenko 1979, Reaves 1974) and large correction factors necessary for the Lick catalogue data mean that statistical quantification of low amplitude (~10%) clustering effects are extremely imprecise and in practice estimates of clustering at large angular scales - corresponding to linear scales in excess of $15h^{-1}$Mpc - are extremely uncertain. An increase in the sophistication of analysis techniques will not improve the precision of clustering estimates at these scales as the data itself is inadequate. Reliable correlation function estimates are severely restricted in linear scale, and not as severe a test of the heirarchical clustering picture as sometimes assumed.

2.4 Detailed Scaling

The scaling of the 2D results and their relation to the intrinsic spatial correlation functions has been treated in detail by Limber (1953), GP77, Phillipps et al. (1978), Fall and Tremaine (1977). In principle once the selection function of an object catalogue is known the spatial clustering may be readily derived from the 2D angular relations. There are considerable limitations on the amplitude of spatial features that may be detected in 2D due to the broad galaxy luminosity function and the large redshift range that contributes to the observed angular functions. Apart from this severe "washing out" of features in 2D, determining the selection function for available catalogues is difficult; reliable galaxy luminosity functions, morphological type frequencies, luminosity evolution and galactic extinction models, for example are not available. Most importantly the selection procedures for the available galaxy catalogues are not simple; the probability of a galaxy's inclusion in an eye selected catalogue depends on surface brightness and profile considerations in addition to the total apparent magnitude (Phillipps et al. 1981), further the image classification probabilities are usually unquantified. As a result, the effective depths and detailed selection functions for the major galaxy catalogues are not well known. GP77, treating the scaling relation between the Zwicky, Lick and Jagellonian catalogues, conclude that the two-point correlation function shows scaling discrepancies of ~20% between Zwicky and Jagellonian catalogues, and for the three-point function the Jagellonian sample exhibits a very large discrepancy ($\geq$50%) compared to the Zwicky and Lick data. Although the discrepancies are not unexpected due to different selection functions for the catalogues, the converse - that the scaling relations for the low order correlation functions are not established to better than these errors - is also

true. Given that a certain degree of optimising the fit between catalogues (by choice of galaxy luminosity function and other factors) has taken place, the scaling relations between the Zwicky and Lick catalogues (only a factor four different in effective depth) are not established to better than 10%. Although the scaling data are consistent with the heirarchical clustering model they do not constitute overwhelming evidence in its favour.

2.5 Linear Scale of Observations

The two-point correlation function estimates from the Lick and Abell catalogues show some evidence of large scale structure up to $150h^{-1}$ Mpc (GP77, Hauser and Peebles 1973, Bahcall and Soneira 1983). Evidence for large scale clustering in the Abell catalogue is stronger than that for the Lick counts, particularly now redshift information has been incorporated in the Abell analysis (Bahcall and Soneira 1983). Evidence for large scale structure in the Lick catalogue is dependent on interpretation; GP77 find large scale structure in one of the four subareas of the Lick catalogue they analyse. GP77 note the presence of the effect and then filter out the large scale features, concentrating their analysis on the smaller scale results. Interpretation of the large scale structure is thus difficult - the possibility of variable galactic extinction or catalogue inhomogeneity can not be ruled out. However, any genuine large scale structure in the galaxy distribution will obviously be removed by the smoothing process. Peebles (1980) states that the results of clustering analyses are unreliable beyond $15h^{-1}$ Mpc but believes there is little evidence for a non-zero covariance function at scales much greater than $15h^{-1}$Mpc. Reliable estimates of the two-point correlation function for galaxies are confined to scales of order $10h^{-1}$ Mpc - compare this to the size of structures postulated by Einasto et

al. (1980) and suggested by nearby redshift surveys (e.g. Gregory and Thompson 1978) of order 50-100h^{-1}Mpc. There exists a large gulf in linear scale between the limitations of the statistical galaxy quantification and the size of "superclusters" indicated from smaller redshift surveys. The statistical results for the clustering of Abell clusters are in better agreement with the ideas of Einasto et al. (1980) and the conclusions from redshift surveys (e.g. Oort 1983) that considerable structure exists on scales of 100h^{-1}Mpc.

The power-spectrum of a portion of the Lick catalogue (including the region found by GP77 to show large scale structure) shows little indication of the strong break from the power-law relation found by GP77 (Fry and Seldner 1982). As already stressed GP77's power-law break is primarily due to the filtering of the large scale structure in the catalogue so Fry and Seldner's result is not suprising. Similarly the two analyses of the Abell catalogue (Hauser and Peebles 1973, Bahcall and Soneira 1983) differ in the amplitude of very large scale structure in the catalogue; Hauser and Peebles taking the view that much of the power at large scales is due to one particularly underdense region of the sky. Bahcall and Soneira's analysis deserves more weight because of the inclusion of redshift information. However, note that the redshift estimates are via Abell's m_{10} index, and 2D variations in catalogue inhomogeneity are not completely independent.

The Abell clustering results of Bahcall and Soneira (1983) give rise to a considerable discrepancy between the results of GP77, Seldner and Peebles (1977) and Bahcall and Soneira (1983). GP77 find the covariance function for single galaxies is essentially zero at scales beyond 20h^{-1} Mpc, Seldner and Peebles (1977) show there is a strong cross-correlation between Abell clusters and single galaxies; approximately 20% of

galaxies being associated with Abell clusters (the majority in cluster halos). The clustering results of Bahcall and Soneira (1983) allow a consistency check between the galaxy and Abell cluster clustering behavior at large scales, assuming that all galaxies not associated with Abell clusters have zero covariance function at large scales, a lower bound to the amplitude of galaxy clustering can be derived by scaling the Abell cluster results by the square of the fraction of galaxies associated with Abell clusters. This fraction is large, suggesting that the galaxy correlation function should be significantly positive at scales of $50h^{-1}$Mpc or more, contradicting the results of GP77. Note that the fraction of galaxies associated with Abell clusters according to Seldner and Peebles (1977) - 20% - is much larger than the 5% adopted by Bahcall and Soneira (1983), making the inconsistency considerably larger than they indicate.

Considering the opposite extreme, attempts have been made to extend the calculation of the covariance function to very small scales (Gott and Turner 1979, Lake and Tremaine 1980). Both invstigations indicate good agreement between the small scale ($<100h^{-1}$ kpc) covariance function behaviour (i.e. power-law slope and amplitude) with the covariance function results at larger scales. These results are somewhat anomalous in a simple clustering heirarchy model: the effects of dissipation, merging and other physical processes become important for galaxy interactions (and hence clustering) at some scale. The simple heirarchical model breaks down at some level otherwise we would not observe such well defined units as galaxies, and there must be some scale at which the observed two-point correlation function deviates from the power-law observed at larger scales. In fact the Lake and Tremaine (1980) results refer to a cross-correlation between high luminosity galaxy systems and very low luminosity galaxy systems. Unless there is a

strong variation in the mass to light ratio of galaxies with luminosity the results do not reflect the mass distribution on small scales, and certainly not the luminosity distribution - this contrasts with the assumption made concerning the large scale results. If very low luminosity systems exhibit a power-law autocorrelation function in agreement with an extrapolation from larger scales in terms of number density of systems, then the clustering behaviour in mass or luminosity terms is very different. An alternative explanation is that the low luminosity systems trace the halo mass distribution of the high luminosity systems, and such systems may be dominated by dissipational effects. In this case the existence of a power-law form of the cross-correlation function at these scales may demonstrate that a number of physical processes give rise to a simple power-law correlation function with slope of order two.

Gott and Turner (1979) consider close bright systems from the Zwicky catalogue: In much catalogue compilation there is a strong tendency to include "interesting" objects which are just outside the normal catalogue selection criteria. No check that both components of the Gott and Turner binaries have magnitudes brighter than m_B=15 has been made, but the magnitude estimates of Turner (1976) add weight to the possibility that significant selection effects are operating in this sample (White and Valdes 1980). There is a clear excess of Turner (1976) binary components with m_B=15 and severe doubts must exist as to whether the Gott and Turner (1979) sample is "well defined". As the observed number of close galaxy pairs increases rapidly for only a small increase in limiting magnitude, evidence for the existence of a power-law correlation function below scales of a few hundred kiloparsecs is poor. This does not directly challenge the simple heirarchical picture as other physical processes are expected to be important at these scales. Perhaps

the most disturbing aspect of the observations is that correlation functions with very similar parameters to those at larger scales are derived from a region where processes other than gravity may be operating.

Two-point correlation function estimates at greater depths than probed by the "standard" galaxy catalogues have been obtained by groups at Durham and Edinburgh using the combination of UKSTU plates and the COSMOS machine (MacGillivray and Dodd 1979, Dodd and MacGillivray 1980, SFEM). These samples total eight plates of six fields, each with an area of 15 square degrees, corresponding to a linear size of $40h^{-1}$ Mpc at the typical sample depth of $Z \sim 0.2 - 0.25$. With suitable scaling and allowance for the somewhat different selection functions involved these should be directly comparable to the earlier Princeton work.

The covariance function estimates from these deeper samples generally follow the familiar power-law relation at small angular scales with the constants A and δ (equation 2.1) in reasonable accord with the shallower samples, although the agreement is not totally compelling (see section 5 of SFEM for a discussion). At large angular scales the shape of $W(\theta)$ shows substantial variations: the break point from the power-law can be at much smaller projected scales, (e.g. $3h^{-1}$Mpc SFEM) or be nonexistent (e.g. MacGillivray and Dodd 1979; Fig. 2a) and significant negative values of $W(\theta)$ are sometimes present (e.g. MacGillivray and Dodd 1979; Fig. 2b, Dodd and MacGillivray 1980; Fig. 2). Substantial discrepancies are present between shallow and deep samples at large angular scales. These are serious as the existence of the power-law break point may indicate the scale at which the change from non linear to linear growth of clustering occurs (Davis et al. 1977) and the failure of the covariance function to become negative at larger scales ($>15h^{-1}$Mpc) would

show that clusters are not surrounded by regions of decreased galaxy density - or "holes" - (Peebles 1974). For a consistent picture of galaxy clustering it is important these discrepancies are resolved. Strong evolution of galaxy clustering can be ruled out because of the short lookback time, and the selection functions for COSMOS samples have been treated in detail (SFEM, Ellis et al. 1977). Chapters 3, 4, 5 and 6 of this thesis consider the problems associated with the UKSTU and COSMOS data used in deep observations, and different methods of determining the correlation function estimates. The most extreme discrepancies evident in MacGillivray and Dodd (1979) and Dodd and MacGillivray (1980) are due to differences in the estimation procedures compared to those employed on the shallow samples. Further disagreements are due to differences in power-law fitting procedures adopted and the limited size of the deep machine measured samples - see Chapter 9.

The failure of GP77 to detect an excess of linear structures in the Lick and Zwicky catalogues when large linear structures are apparently evident to the eye has led to a general acceptance that correlation functions are insensitive to the presence of linear features (e.g. Groth 1980). Examination of the three-point correlation results (GP77, Fry and Peebles 1980) shows that the scales actually examined are small. Further, the presentation of the results by averaging over the variables "θ", "U" and "V" (e.g. GP77) is not the optimal method of enhancing the signal to noise of the statistic to detect filaments of finite width - see Chapter 9. The effective linear scale of the calculated three-point function is less than the maximum separation of $15h^{-1}$ Mpc as the averages are dominated by the contributions from many close triplets. GP77 examined the three-point function in the Lick catalogue to scales of five degrees in an attempt to detect linear structures on scales of 0.5-3.0 degrees. Eye examination of the Lick counts suggests that

prominant linear structures are present on scales of several tens of degrees with their smallest dimensions subtending three degrees on the sky. These features are comparable in linear scale to the structures postulated by Joeveer et al. (1978), the individual features studied by Joeveer et al. (1978), Einasto et al. (1980), Chincarini et al (1981) are also of order $10h^{-1}$Mpc in their smallest dimension. Again as with the two-point function there is a large discrepancy between the size of structures postulated to exist from individual object studies and the scale of the statistical quantification available - this point is not generally appreciated. Similar comments apply to the analysis of the cluster-galaxy-galaxy correlation function examined by Fry and Peebles (1980). The maximum scale examined is only $15h^{-1}$Mpc and the result relating to the mean ellipticity of Abell clusters (i.e. mean ellipticity is 0.65 ± 0.15) may be interpreted differently by those favouring the existence of linear structures.

It is certainly true that the three-point correlation function is not the most sensitive method for detecting linear structures, but to date the statistic has not been applied at scales on which apparently obvious filamentary structure is present - the criticism of the ability of the three-point function to detect linear structure is not yet well founded. Conversely the range of linear scales for which reliable estimates of the higher order correlation functions are available is small, and the statistical analyses have not probed the linear scales where differences between the heirarchical and adiabatic collapse theories may be expected. Once again the evidence favouring the heirarchical model is not overwhelming. The complexities of dissipational processes do not allow even a qualitative prediction of the form of the higher order functions at small scales to be made, but this is hardly a strong arguement against the adiabatic collapse model.

2.6 Redshift Surveys

From the mid-1970s sufficient galaxy redshifts have been available for direct analysis of the spatial distribution of galaxies. Several types of surveys have been completed; nearby (Deff ~ $50h^{-1}$ Mpc) wide angle surveys (Yahil et al. 1980, Fisher and Tully 1980, Davis et al. 1982), deeper narrow angle surveys of field regions (Kirshner et al. 1978, 1979, Bean et al. 1983; in preparation), and narrow angle surveys towards specific galaxy groupings (e.g. Gregory and Thompson 1978, Gregory et al. 1981, Williams and Kerr 1981, Tarenghi et al. 1979, 1980, Chincarini et al. 1981).

The nearby surveys have mapped and delineated the structure of the local supercluster (LSC) to high precision - a review is given by Tully (1982). The data for the LSC are now the most complete for any large galaxy system - only in the LSC for instance is data on the distribution of low luminosity galaxies available. Detailed descriptions are given by Tully (1982) and Yahil et al. (1980), from which the following points are of interest; (a) evidence for a highly flattened system exists with a major diameter of 10-15 Mpc and axial ratio of order 6:1, (b) a number of smaller flattened groupings of galaxies appear to be associated with this major plane, (c) the filling factor of the major plane and smaller clusters is extremely small; statistics relating to filling factors can be deceptive but Tully's data show approximately 80% of the galaxies within the LSC are located in discrete groups that fill 10% of the volume. Huchra and Geller (1982) find 60% of the galaxies in a subset of the Harvard-Smithsonian survey occupy 5% of the survey volume. In addition the number of isolated field galaxies is small; Huchra and Geller's figure of 26% is an upper limit, a large fraction are expected

to be due to effects of the survey boundaries and the magnitude selection criteria of the sample. (d) The radial number density fall-off of bright galaxies from the center of the LSC follows an inverse square relation. Note, that the size of the system combined with the radially averaged number density fall-off is consistent with the two-point correlation function results described earlier, (e) the LSC is an order of magnitude smaller than many of the systems found in studies of more distant regions (e.g. Einasto et al. 1980, Gregory and Thompson 1978).

The Harvard-Smithsonian galaxy redshift survey (Davis et al. 1982) contains data extending beyond the LSC; redshifts for 2400 galaxies survey the galaxy distribution to $\sim 80h^{-1}$ Mpc. The sample consists of galaxies brighter than $m_B=14.5$ in the Zwicky catalogue. The survey is unique in providing data over a large solid angle to reasonable depth.

A detailed statistical analysis of the clustering in the catalogue is not complete and will be difficult due to the small number of objects, Zwicky-magnitude selection effects and strongly varying redshift selection function (Davis and Huchra 1982). A visual inspection of the figures in (Davis et al. 1982) shows a clumpy distribution of objects with structures of $10-20h^{-1}$Mpc clearly visible, a few larger structures are visible (e.g. Coma/A1367) and regions of $20h^{-1}$Mpc virtually devoid of galaxies are evident. It should be stressed that the occurrence of galaxies in clusters greatly reduces the number of independent structures in the diagrams; the presence of large voids is a natural consequence of disributing a small number of random points in a large volume. Some prominant "filamentary" structures are visible but there are also numerous spherical structures, and the frequency of occurrence of filamentary structures has yet to be shown to be in excess of that expected. Supercluster and void sizes in the Harvard-Smithsonian survey

are generally at the lower limit of the structure scales suggested from other studies (e.g. Einasto et al. 1980, Thompson et al. 1981). Oort (1983) reviewing the Harvard-Smithsonian survey estimates the size of the Coma/Al367 structure as up to 160*24 Mpc, but the data shows little evidence for the structure extending beyond $40h^{-1}$ Mpc in its largest dimension - comparable in size to two adjacent cluster/halo complexes of the type found by Seldner and Peebles (1977). The consequences of the random clustering of large ($10-20h^{-1}$ Mpc) structures have not been fully enumerated.

3D surveys of small areas of sky toward specific objects provide more detailed information over much reduced volumes. A number of points concerning selection procedures in these samples are; (a) regions are chosen where there is a large aglomeration of galaxies of comparable magnitude - usually away from confused areas of clustering, (b) the magnitude limit of the galaxy sample is bound to peak close to the magnitude of a typical galaxy in the structure, resulting in few objects being visible behind the structure, (c) the foreground volumes surveyed are small, as are the transverse extents of the surveys in megaparsecs. Plots of the resulting redshift distribution in "expanded cone diagrams" can be highly misleading; Fig. 3 of Tarenghi et al. (1979) appears extremely impressive in terms of the size of the "voids" present in front of Hercules. The same data combined with additional redshifts, plotted correctly scaled (Tarenghi et al. 1980; Fig. 3) gives a much fairer picture of the galaxy distribution - the scale of the "voids" is much reduced. Similar comments apply to the cone diagrams of Gregory and Thompson (1978; Fig. 2) toward Coma.

The scale of voids ($\sim 50h^{-1}$ Mpc) and superclusters together with associated high galaxy density contrasts (factors of order a hundred) derived from

this type of survey (e.g. Gregory et al. 1981) have yet to be shown to be completely valid. Confirmation requires deeper surveys combined with greater angular scale, and a more objective choice of areas for study.

In contrast to surveys toward known clusters, attempts have been made to survey "field" regions with no obvious structure (Kirshner et al. 1979, Shanks 1982, Kirshner et al. 1981). These surveys are confined to small numbers of galaxies ($\leq$300) and very restricted solid angles - only the redshift ordinate contains information on clustering scales exceeding $10h^{-1}$ Mpc. Combination of the strongly varying number-redshift relation and the small scale clustering of galaxies makes analysis of these surveys difficult. The problem of determining the expected number-redshift relation via data on the galaxy luminosity function and other parameters is illustrated in Davis and Huchra (1982); fluctuations in the number-redshift relation are evident on scales of $30h^{-1}$ Mpc. The narrow angle surveys are far more susceptible to fluctuations of this type due to the presence/absence of only one cluster in the line of sight. Kirshner et al. (1979) found an excess of clustering on scales $10\text{-}30h^{-1}$Mpc compared to the results derived from 2D observations by Peebles, but note the significant dependence of the large scale structure on the adopted luminosity function (Kirshner et al. 1979; Fig. 7). A reanalysis of this data combined with the Durham group's data (Shanks 1982) gives completely different results - a power-law dependence of $\xi(r)$ holding to $5h^{-1}$ Mpc with the detection of negative values of $\xi(r)$ on scales $15\text{-}50h^{-1}$Mpc. This dramatic change in results from the same data emphasises the care that must be taken with deep narrow angle surveys.

The lack of appreciation of simple selection problems is illustrated by Kirshner et al. (1981); surveys of three narrow cones (width $4h^{-1}$Mpc at

the peak of the theoretical number-redshift relation) revealed a coincident gap in redshift space. An extrapolation of more than two orders of magnitude in surface area was made to postulate the existence of a large void(!) - the number of clusters in each field is small and subsequent analysis of more redshifts has reduced the size of the void considerably (Balzano and Weedman 1982, Sanduleak and Pesch 1982).

2.7 Summary

The deep narrow angle redshift surveys combined with the larger Harvard-Smithsonian data suggest the existence of superclusters and voids at the smaller end of the size spectrum postulated by Einasto et al. (1980), Tarenghi et al. (1980) and others. The observations of voids has stimulated some theoretical work (Aarseth and Saslaw 1982) showing that void sizes up to $30h^{-1}$Mpc can be expected in the simple heirarchical clustering model. Further theoretical attention to the possibility of significant changes in the mass to light ratio of matter in the universe on these scales is required. This problem recalls the assumption that bright galaxies are a good tracer of the mass distribution in the universe. In practice there is little good evidence to demonstrate that even the galaxy luminosity function is invariant with position for systems more than two magnitudes below the knee of the Schechter luminosity function (e.g. Felton 1977).

Einasto's work (Joeveer et al. 1978, Einasto et al. 1980) identifying cellular structures in the galaxy distribution remains impressive, however, earlier remarks concerning the small number of independent clusters concerned apply. More redshifts are required, and the question remains of whether there is a statistically significant number of ring or cellular structures in excess of that expected - given galaxies are

grouped into clusters of dimensions $10-20h^{-1}$ Mpc.

Although the evidence for highly asymmetric superclusters is not entirely conclusive the limits on the occurrence of asymmetric structures imposed by the correlation results are weak, primarily due to the small range of scales examined. Evidence is also growing that power-law correlation functions at small scales are a consequence of other galaxy formation scenarios (e.g. Melott 1983; preprint).

As a result of small samples, poorly determined galaxy catalogue selection effects and the limited range of linear scales for which a quantitative statistical description of galaxy clustering is available, the distribution of galaxies is poorly determined. The observations to enable discrimination between theoretical models of galaxy formation (e.g. Peebles 1980, Zeldovich 1978) have yet to be made.

3 The COSMOS Machine

3.1 Introduction

The availability of the COSMOS machine for measurement of large numbers of wide field photographic plates, combined with the apparently high quality of data being produced at the start of the investigation was the primary reason a large scale clustering project was considered feasible. Initial investigation of COSMOS measures of Anglo-Australian Telescope (AAT) and UKSTU plates revealed considerable problems in the data - positional dependent success rates of image classification for instance. Discussion with Image and Data Processing Unit (IDPU) and UKSTU members, as well as a literature search, revealed little quantitative information on possible causes of the systematic effects. In several cases it was not obvious whether the plates or the measuring machine were responsible, and a parallel investigation of UKSTU plate properties was undertaken - Chapter 4.

In this chapter a description of the relevant features of the COSMOS measurement process are given, followed by an account of the problems encountered in the COSMOS measures. Two areas were singled out for particular attention; the COSMOS calibration procedures, and the spot characteristics. First, an outline of some of the practical difficulties encountered in the interaction with the COSMOS machine is given. These difficulties made progress in the investigation very slow, preventing to a large extent the completion of the project as originally envisaged and confining the available data to a number of single UKSTU plates.

3.2 Practical Difficulties with the COSMOS Investigation

Several factors combined to prevent an understanding of COSMOS measures of UKSTU plates until very late in the project - early 1982; (a) an almost total lack of documentation relating to the overall performance of the COSMOS machine, and to individual COSMOS measures of plates - this continued throughout the project, (b) no specific COSMOS performance specification was available against which results could be checked and quality control performed, (c) the competition from many other projects for machine time meant the time scale between requests and receipt of test data was often long (up to six months), and follow up data to test anomalous results was not always available, (d) the development of the machine over the period of the investigation resulted in major hardware and software analysis modifications which changed the machine performance significantly at irregular but frequent intervals. Little comprehensive testing was performed to ascertain the nature of the performance changes.

As a result, considerable time was spent analysing a large number of COSMOS measures to ascertain performance "norms" and search for correlations with many possible factors, and all test and diagnostic software had to be written as part of the investigation. Table 3.1 outlines the main changes and periods of downtime for the COSMOS configuration during the investigation period. Progress was also affected by the lack of quality control which resulted in many plate, and several test measures being lost due to one-off "machine glitches". In effect, a program to determine a detailed COSMOS specification was carried out, with software techniques developed to define performance characteristics and the amplitude and nature of field effects in the data. Early in 1982 a set of specific quality control procedures were

TABLE 3.1

Date	Description of change
September 1980	Installation of autofocus
January 1981	Final acceptance tests for autofocus
February-November 1981	Discovery and rectification of faults in original background filtering procedure
March 1981	New cathode ray tube (CRT)
March 1981	COSMOS down for CRT and hardware installation
March-April 1981	Severe stripey data problem
April 1981	Default I.A.M. increment changed from 8 to 16 microns
June 1981	Transmission resolution of I.A.M. data increased by a factor two
July-December 1981	Coordinate carriage "sticking" and spurious image generation (double vision)
December 1981	COSMOS down for rectification of "sticking" problem
December 1981	Installation of median background filter
December 1981	Default value of Tb changed from 0 to -6
January 1982	Faults in median filter rectified
January-April 1982	No background filtering performed
March 1982	New CRT
May 1982	Median filtering correctly implemented
May 1982	New CRT

submitted to the IDPU group, a large number of these were implemented - resulting in immediate improvements in the quality of data supplied to users.

3.3 COSMOS

A full description of the COSMOS machine and the associated reduction software will not be given here, details may be found in IDPU

documentation (COSMOS handbook 1982). The following factors were relevant to the investigation at various stages, and brief descriptions of each are given.

3.3.1 Scanning and Detector Systems; The COSMOS scanning light source is a Ferranti microspot cathode ray tube, with nominal spot sizes (focussed on the plate emulsion), of 8, 16 and 32 microns full width at half height (FWHH) - the 32 micron spot is used for virtually all COSMOS scans. A large halo of unspecified size is associated with light sources of this type (see Section 3.7), limiting measurement of high densities. The light transmitted by the photographic plate is collected via a light pipe and registered by a photomultiplier. A second beam from the light source is used in a divider loop to remove transmission fluctuations due to variations in the intensity of the incident beam - but see the conclusions on COSMOS stability in Section 3.6.3.

The recorded light transmission (T) for each pixel is digitised to 255 levels $1 \leq T \leq 255$. The zero point is adjustable so that the limited dynamic range may be optimally utilised - within the constraints applying at high density due to the spot halo. An autofocus system was installed (about a year into this project) which keeps the very small depth of focus for the spot - 10 microns - on the emulsion (an original UKSTU plate may sag by 250 microns from edge to center when mounted in the COSMOS machine), this was a potentially important change in machine performance.

The electronic systems incorporate a number of different gain settings that allow the signal from the photomultiplier to be reduced by fixed factors before the digitisation to output T values occurs. The different gain settings allow high light transmissions to be recorded without

transmission saturation occurring - i.e. T>255. This effective increase in dynamic range is used for plate calibrations, (Section 3.6) but measurements of plates are performed at a single gain. The nominal dynamic range of the gain 1 setting - the usual measurement gain - is D ~ 2.0.

3.3.2 Scanning Mechanics; Plates are mounted horizontally in a plateholder on a movable plate carriage. The spot light source and detector system are fixed in position and mounted above and below the plate carriage respectively. The spot performs raster scans of 128 pixel increments parallel to the COSMOS X axis. Increment sizes available are 8, 16 and 32 microns corresponding to raster lengths of 1024, 2048 and 4096 microns - one raster length defines a lane width. As the spot scans in X the plate carriage is moved in Y over the extent of plate to be measured. The rate of drift is such that successive raster scans abut, the maximum Y range is approximately 280mm. The movement of the carriage through the Y range to be measured completes the scanning for one lane. The process is repeated (the carriage incremented by one lane width in X), until the area of plate to be measured has been scanned - the maximum X range is approximately 260mm. The nature of the raster scan coupled with the fixed source/detector configuration results in a different part of the microspot emitting the light recorded by each pixel in the scan. For periods when the microspot phosphor response was particularly uneven, artifacts were introduced into the COSMOS I.A.M. scans. These manifested themselves as modulations in the frequency of images detected, and their properties, as a function of position within a lane. A number of COSMOS measures for this project were severely effected by this problem. The modulation in image detection across a lane is often visible in COSMOS measures.

3.3.3 Measurement Modes; Two main measurement modes are available. Mapping Mode which sequentially records the T values for every increment in the scan to be measured - subsequent offline analysis is then performed by the user. This mode was rarely used for large areas of plate due to the huge quantity of data - more than one hundred magnetic tapes for a Schmidt plate digitised at 8 micron increment, although since early 1982 mapping of plates at 16 micron increment with offline smoothing has been implemented.

Image Analyses Mode (I.A.M.) is most frequently used, and is specifically designed for the analysis of large areas of photographic plates. The first stage is the calibration procedure (see Section 3.6) to provide a lookup table (within the COSMOS machine), between recorded T values and relative intensity. COSMOS then determines a sky-background value by calculating the median of the T value histogram from a sample area one lane width on a side - e.g. 1024 microns for 8 microns increment, every sixteenth pixel in a square pattern is sampled to produce the histogram. The raw sky values are then filtered (Section 3.3.5) and the resulting values of sky-intensity (per square, one lane width on a side) are allocated to every image falling within the area boundaries - no interpolation between the centers of sky blocks is performed. A threshold at a fixed percentage of the adopted sky value in intensity space is then applied to all the pixels in the block. The threshold is set at the start of each measure and only pixels with intensities greater than this threshold are output to a magnetic tape.

Subsequent analysis of the thresholded data is performed offline. The detailed procedure is not important; basically algorithms to identify numbers of contiguous pixels are implemented. A lower image area cut is applied to remove spurious images due to emulsion or machine noise, and

Table 3.2

Parameter number	Parameter description	Parameter units
1	X centroid (unit weighted)	0.1 microns
2	Y centroid (unit weighted)	0.1 microns
3	X minimum	0.1 microns
4	X maximum	0.1 microns
5	Y minimum	0.1 microns
6	Y maximum	0.1 microns
7	Number of image pixels	Increments
8	Minimum Transmission	T levels
9	-250*Log(Σ(I - Isky))	Intensity
10	Sky intensity at image centroid	Intensity
11	X centroid (intensity weighted)	0.1 microns
12	Y centroid (intensity weighted)	0.1 microns
13	Semi-major axis Au (unit weighted)	0.1 microns
14	Semi-minor axis Bu (unit weighted)	0.1 microns
15	Orientation of major axis (unit weighted)	degrees
16	Semi-major axis Ai (intensity weighted)	0.1 microns
17	Semi-minor axis Bi (intensity weighted)	0.1 microns
18	Orientation of major axis (intensity weighted)	degrees

(a) The minimum transmission can be converted to a maximum intensity via a lookup table in the tape house-keeping block. A "peak surface brightness" can be defined as Log((Imax - Isky)/ Isky), henceforth denoted PSB.

(b) Parameters "9" and "10" can be combined to give a thresholded magnitude above sky by converting parameter "10" to intensity per square arcsecond (Iarcsec); the quantity -2.5 * Log(Σ(I-Isky)/ Iarcsec) defines "cosmos-magnitude".

(c) An effective radius R^* may be defined as sqrt(area/π) and calculated from parameter "7".

simple moment analysis techniques are applied (Stobie 1980), the object parameters listed in Table 3.2 are output to tape.

3.3.4 COSMOS Background Determination; from the raw sky-intensity (determined as described in the previous section) a final array of sky-intensities is derived using a filtering technique. Until mid-1981 the filtering technique applied resembled a Laplacian operator (Stobie 1981, IDPU internal report), which identified regions where the second derivative of the background was large. The operator requires a scale "k" and threshold level for operation. If the magnitude of the operator is greater than the threshold level the background value is replaced by the mean of the background pixels distance $\pm k$ away, otherwise the background value remains unchanged.

Two problems arose; (a) with the filter concept, (b) with its implementation on COSMOS. (a) The filtering is performed only in the direction of the COSMOS Y axis and is thus one-dimensional, further, Stobie (1981, IDPU internal report) showed that the filter propagates sharp features and large gradients in the background a considerable distance into originally unaffected background - the edges of stepwedges or occurrence of bright stars contaminated large areas due to the filtering. The practical effects of using the filter with a large scale length can be seen in Fig. 2 of Shanks et al. (1980a). Additional problems occurred with the use of different filter lengths and threshold values at various times, and data changed significantly depending on the exact values used - some combinations produced highly anomalous backgrounds.

(b) In 1981 the implementation of the filter in the COSMOS machine was found by IDPU members to contain numerous coding errors, the code for

the background following was implemented in computer assembler language which made recognising errors extremely difficult. A particular problem with the original filter was the propogation of incorrect sky-intensity boundary values into the final data. These values are input at the start of a measure to provide a reference point, enabling the filter to "get started" from the corner of a measure. This resulted in several measures being discarded. In some periods no filtering was performed at all, though this was not discovered at the time. It was not clear what the implementation of the filter was doing under certain conditions and several unexplained anomalies in COSMOS measures were probably due to errors in background filtering. The adoption of a new median filtering technique in December 1981, and correct implementation in May 1982 contributed significantly to the improvement in COSMOS data.

3.4 Thresholded Data

It is important to appreciate the limitations and effects of thresholded data: the thresholding philosophy means all photometric parameters are defined relative to the threshold base - in this case the local value of the sky-background. To obtain the correct thresholded magnitude for an object, the COSMOS parameter $2.5*\mathrm{Log}(\Sigma(I - Isky))$ must be divided by the sky-intensity. If this is done for the sky-intensity per square arcsecond we have a convenient measure of thresholded magnitude above "sky". For example a value of -2 indicates an object has a thresholded magnitude two magnitudes brighter than sky. Approximate zero points can be deduced by adding the sky brightness per square arcsecond - typically 22.5 at Siding Spring in the J band - giving the example a thresholded magnitude of 20.5. All references to cosmos-magnitude refer to this thresholded magnitude above sky and should not be confused with total magnitude. As objects become fainter a smaller fraction of the total

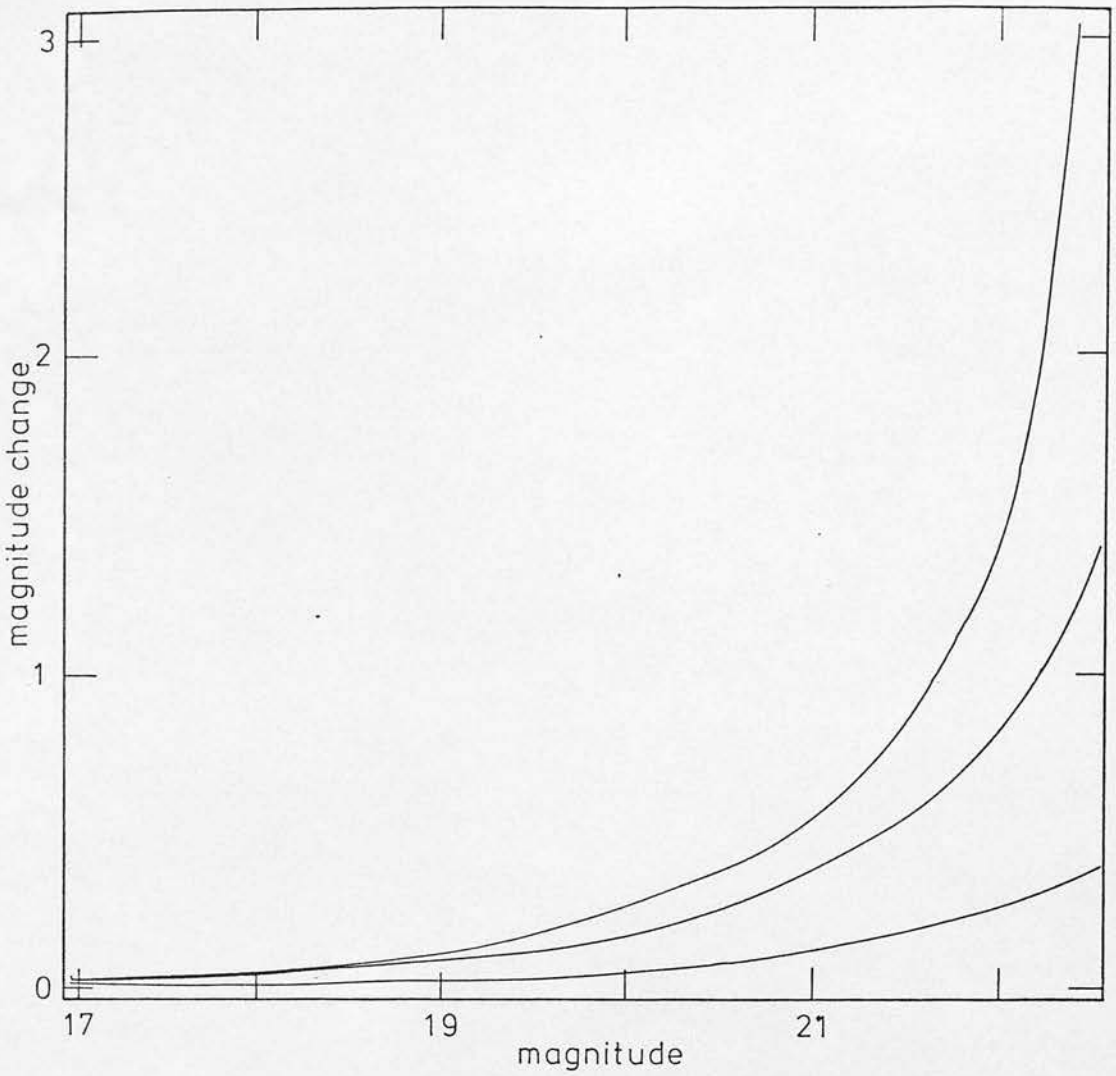


Fig. 3.1 The magnitude reduction as a function of magnitude for three images profiles due to the COSMOS thresholding procedure. The images are idealised profiles (Moffat 1969) and have been chosen to approximate a stellar image with typical seeing (1.8 arcseconds) - lower curve, a typical galaxy - middle curve, and a very low surface brightness object. The radii containing 90% of the total light for the latter images being 3.7 and 7.4 arcseconds. The sky-brightness is 22.5 magnitudes per square arcsecond, and the threshold cut 10%.

intensity of the object is included in a thresholded magnitude. Consequently cosmos-magnitude is systematically too faint when compared to a total magnitude estimate. The effects of this, on object detection, magnitudes and colours can not be overemphasised. Fig. 3.1 illustrates the size of the effect for idealised profiles (Moffat 1969) in intensity space. The threshold level is 10% of the sky-background which is 22.5 magnitudes per square arcsecond. The effect is most severe for galaxies due to the more diffuse profiles. The effects of thresholding must be incorporated in any model of the COSMOS measurement process.

A property of the thresholding technique is the invariance of the photometric parameters to the amount of light reaching the plate, or to changes in the probability that a number of incident photons will give rise to a developed grain. Two such effects are telescopic vignetting - decrease of incident intensity with radius from optic axis - and emulsion senstivity changes. As both phenomena effect the sky and object intensities in fixed proportion, the thresholding relative to sky maintains the relation between sky and object; COSMOS I.A.M. parameters should not be effected by vignetting or emulsion sensitivity changes.

The properties of thresholded magnitudes in the presence of background variations were not universally appreciated at the start of this investigation. Another "magnitude" system was in use; direct application of the COSMOS summed intensity parameter without sky normalisation. Combined with the uncertainties over the cause of the sky-background variations this resulted in some confusion. Clearly, failure to normalise by sky-intensity in the presence of vignetting or sensitivity variations will result in apparent "field effects" in the COSMOS photometric parameters. In defense of this approach, for the presence of additive sky-background variations division by local sky is not the

correct procedure. In all cases however the COSMOS threshold level is determined as a percentage of sky, and will be incorrectly set when additive variations are present - the COSMOS measuring procedure can not correctly take additive sky-background effects into account. Possible causes of additive sky-background changes include; chemical fog variations, low surface brightness nebulosity, zodiacal light and halation surrounding bright stars due to reflected light. In practice both types of effect will be present to some degree. The presence of 10% background changes due to either cause coupled with an incorrect measurement procedure will result in a COSMOS magnitude change of > 0.1 magnitudes. Table 3.3 gives examples of the amplitude of errors in cosmos-magnitude due to the presence of background variations and sky-background misdetermination.

Table 3.3

Magnitude	Additive		Misdetermination	
	+3%	+5%	+1%	-1%
19.5	0.04	0.06	0.03	-0.03
20.0	0.04	0.07	0.04	-0.04
20.5	0.05	0.07	0.04	-0.04
21.0	0.05	0.09	0.06	-0.06
21.5	0.09	0.13	0.10	-0.10
22.0	0.34	0.51	0.22	-0.19

(a) calculations for Moffat profiles (R=2.5, beta=3.0), sky-brightness 22.5 magnitudes per square arcsecond and threshold 10%.

The discussion has assumed that the correct density to intensity calibration has been applied to the plates, all the above remarks relate to data in intensity space. Further discussion of this point is contained in Section 3.6.

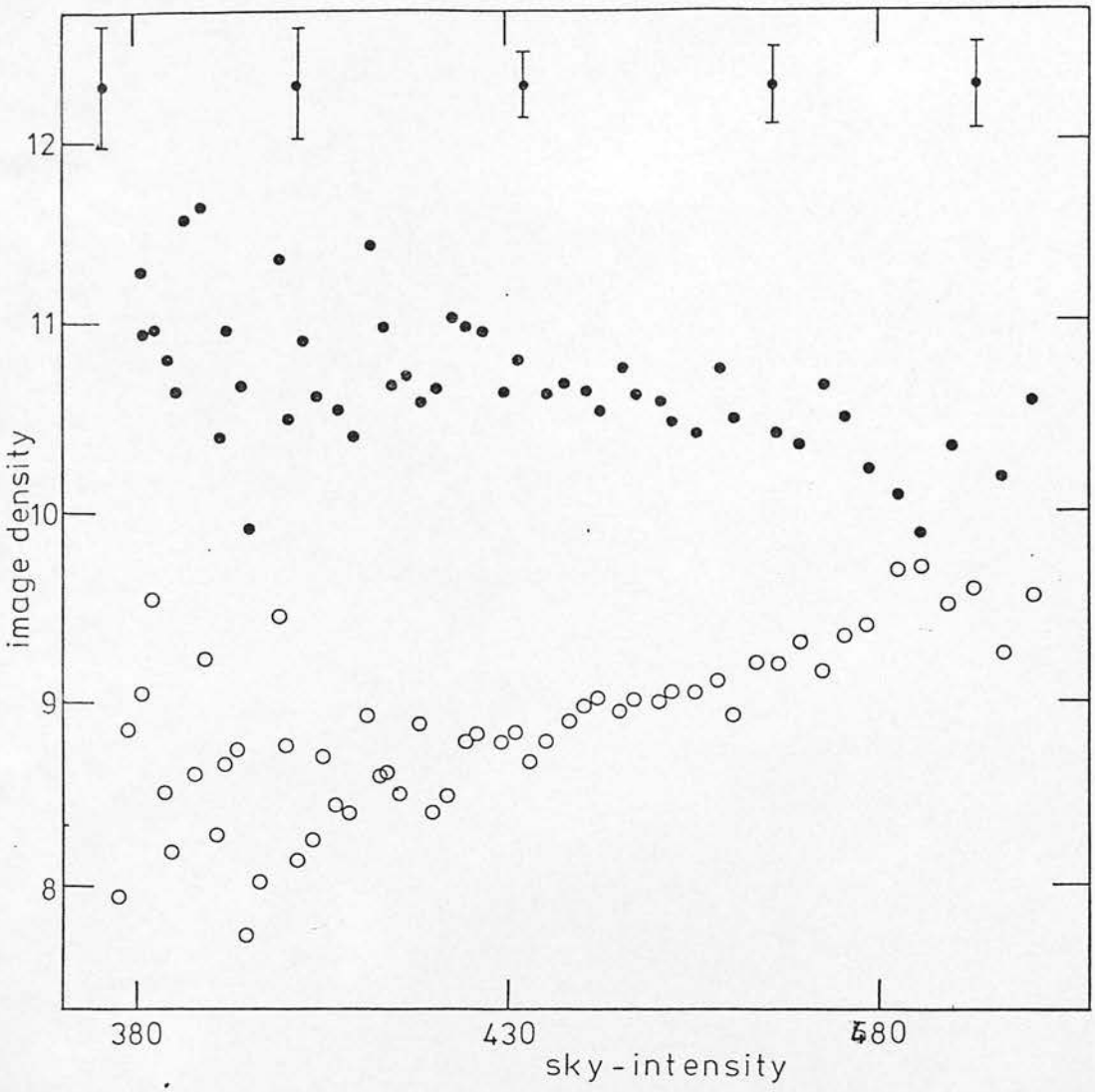


Fig. 3.2 The surface density of "stars" (O) and "galaxies" (●) as a function of COSMOS sky-intensity on plate J3721. Images were classified using the technique of MacGillivray (1976). The density axis is relative and the curves have been shifted vertically for clarity.

3.5 Preliminary COSMOS Results and Parameter Shifts

Analysis of UKSTU and AAT plates in the first half of 1980 using image classification software developed at ROE (MacGillivray 1976) showed large systematic changes in classifier success rates as a function of plate position. The ratio of stars to galaxies on UKSTU plates appeared to vary by factors of three or more. The systematic changes showed a broadly radial dependence on plate position.

Fig. 3.2 shows the number of stars and galaxies classified on a COSMOS measure of plate J3721 as a function of radial distance from the plate center. An apparently related problem was the large change in image number density with plate position. IDPU members were generally of the opinion that many systematic changes in the data with plate position were due to defocussing of the COSMOS spot on the emulsion because of plate sag (section 3.3.1), and little attention was paid to the effects as an autofocus system was to be installed. After the installation of the autofocus system in October 1980, examination of several COSMOS measures revealed that systematic changes were still evident, although some of the gross image number density variations had been eliminated. Changes in "galaxy" density of 100% were still present. Some doubts were expressed as to the magnitude of the effects but four explanations were proposed; (a) autofocus was not working correctly, and some defocussing was still present, (b) instabilities in the COSMOS electronics over the period of a scan resulted in systematic changes in the transmission values recorded, and hence image parameters with plate position, (c) the sky-background variations were additive in nature and the problem lay with the thresholding technique, and (d) the plates possessed large intrinsic field effects of an unspecified nature, and the variation in

parameters had nothing to do with the measurement by COSMOS.

As no data was available on possible causes of the problem, a large quantity of data was built up from COSMOS measures; a total of thirty plates with COSMOS measures were examined. A wide range of factors were investigated, variations with emulsion type, plate quality, copy plates, background density and as many different settings of the COSMOS measuring procedure as possible. Unfortunately multiple COSMOS measures of single plates were infrequent at this time. The following conclusions emerged from this study; (a) one type of systematic effect, confined to IIIaJ and IIIaF plates, was strongly correlated with the local COSMOS sky-background intensity, but numerous peculiarities were present on all plate types. These varied from plate to plate and appeared to be different in nature from the systematic background correlated effects on the type III emulsion plates. The absence of the systematic effects on type II emulsions ruled out the autofocus system as a cause. This was confirmed by the presence of systematic effects on copy plates which exhibit much reduced sag due to their increased thickness. COSMOS measures of plates with and without autofocus however did reveal considerable improvements in data quality with autofocus present.

(b) The possibility of transmission drift due to instabilities within COSMOS (Stobie 1981, IDPU internal report) was not fully discounted until the Joyce Loebl measures described in section 3.6 were analysed. However the absence of the systematic effects in measures of type II emulsions required a considerable coincidence between the occurrence of the effect and the measurement of type III emusion plates. Further the form of the variations; the correlation with sky-background and broadly radial dependence did not fit the T drift hypothesis. A time drift effect would be strongly correlated with COSMOS X coordinate.

(c) The sky-background correlated systematic variations in parameters were present on sky-limited AAT plates, where the sky-background variations were due to vignetting. There should have been no correlation with sky-background on these plates if the COSMOS measuring procedure was operating correctly. This ruled out the possibility that the effects were entirely due to additive sky-background variations.

(d) There was some correlation between plate quality - determined from the UKSTU gradings - and the apparent quality of the COSMOS data, but the correlation was one with the amplitude of random errors rather than the presence of systematic effects. Only in extreme cases where additive background changes were present, (due to processing problems or defocussing in copy plates) were systematic shifts in image parameters correlated with features visible on the plate - excluding the sky-background correlation already discussed. No correlation with changes in image structure over the plates was found. Of the few COSMOS multiple measures of plates, at least two pairs (of plate B3499) showed very large systematic differences in the COSMOS parameters as a function of plate position. This confirmed that some types of systematic shift in the COSMOS data were entirely attributable to the machine. However, many UKSTU plates measured by COSMOS are not of the highest quality, and show processing defects that modify the background by up to 0.25D. In cases where background changes due to processing are of this order the plates themselves contribute significantly to the observed field effects.

The existence of several types of errors in the measures made quantification and identification of the causal factors time consuming. The occasional one-off type of glitch affecting only a few measures was particularly difficult to identify.

The correlation of the systematic changes in image parameters with the sky-background prompted a more detailed analysis of the nature of these systematic effects. Ten COSMOS measures of direct sky-limited IIIaJ and IIIaF plates were examined further. As one of the most obvious manifestations of the systematics were shifts in the region occupied by images in several COSMOS I.A.M. parameter spaces, the systematics will be termed "parameter shifts".

Two methods of quantifying the parameter shifts were attempted; (a) selection of image samples in a narrow range of one parameter and examination of the change in the values of a second as a function of other factors, (b) the quantification of the change in position of identifiable features in parameter spaces as a function of other factors.

Checks were performed to examine the magnitude of the parameter shifts defined in this fashion as a function of photometric COSMOS I.A.M. parameters. In all cases the change in parameters was strongly correlated with the local value of sky-intensity. The sense of the correlation being that objects became brighter at high sky-intensity - an effect opposite to that expected for additive sky variations. No other parameter showed significant correlations that couldn't be explained by the sky-intensity effect. Note that COSMOS saturation at high densities results in an effect opposite in sense to that observed - peak surface brightness decreases with sky-intensity. Similar behaviour was evident in most measures, though the percentage change in the parameters varied considerably from plate to plate, and the variations were not always simple linear functions of sky-intensity. A possible cause of the effect is a variable number of galaxies in each sky range,

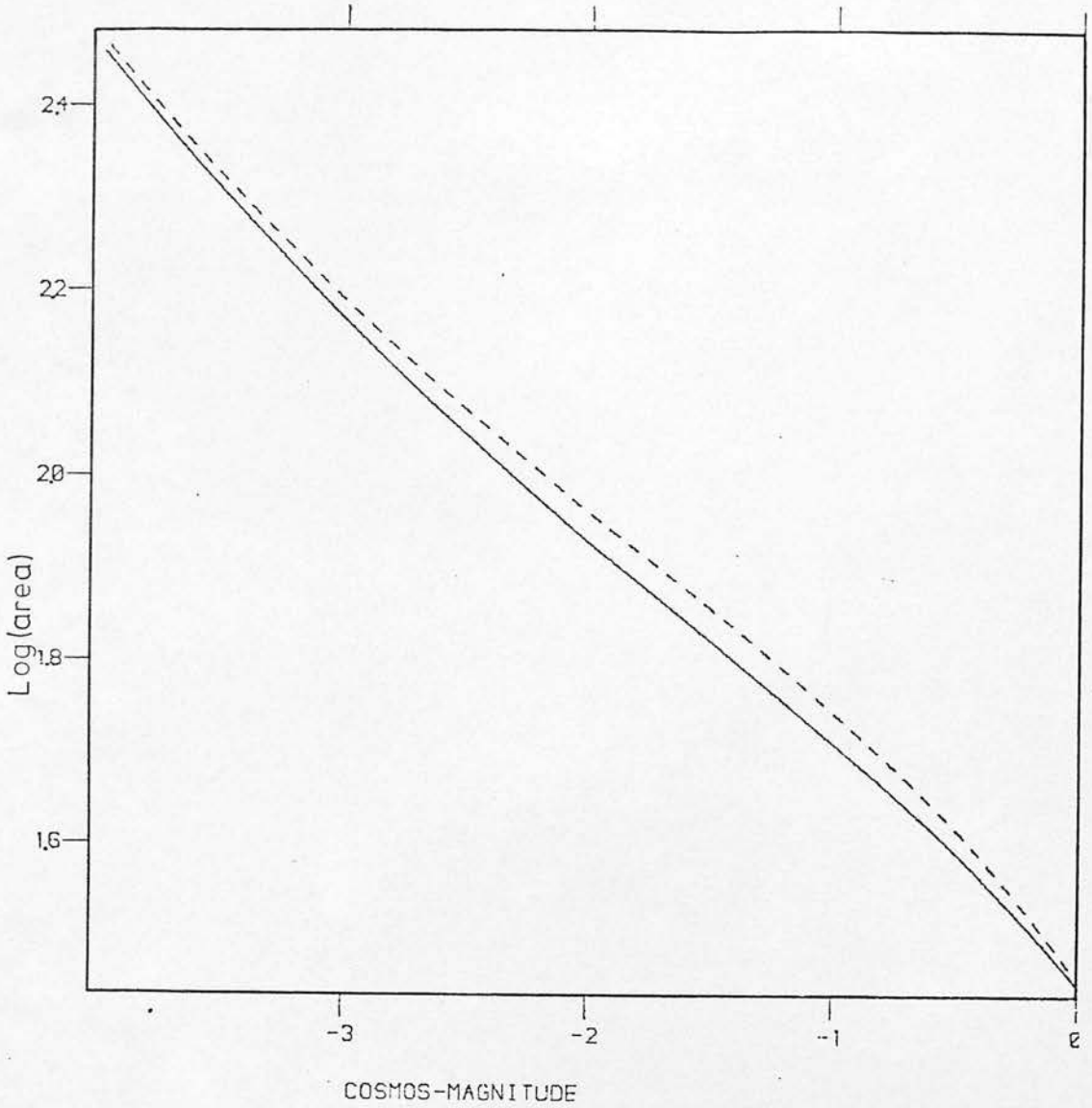


Fig. 3.3 The shift in the position of the stellar locus in the cosmos-magnitude vs. Log(area) parameter space for plate J2652. The dashed line is for images in the lower 50% of sky-background, the solid line for images in the upper 50% of sky-background. The maximum distance between the loci in terms of the local gaussian dispersion is 1.6 sigma.

but the data showed the same behaviour when confined to stellar objects, and eye examination of plates ruled out variable galaxy number densities as a cause.

For data to be useable over wide areas of plate the position of identical objects must be invariant in a parameter space regardless of object position or other factors. A group of images that lie in a narrow sequence in many COSMOS I.A.M. parameter spaces are single stars, an individual stars' position relative to the sequence locus is determined only by random errors - neglecting minor colour and telescopic effects. The sequence of stellar objects in most COSMOS I.A.M. parameter spaces resembles a gaussian in cross-section. A measure of the shift of images in a parameter space is the distance in terms of the local gaussian dispersion that the sequence locus moves.

The position of the stellar locus and calculation of sigma at each position along the sequence were determined using non-linear least squares techniques to fit gaussian number density profiles to cross-sections of the stellar sequence. Fig. 3.3 shows an example of the shift of the stellar locus for images on plate J2652 in the upper and lower halves of the plate sky-background. The peak amplitude of the shift exceeds 1.6 times the local gaussian dispersion. The amplitude of the effects varies considerably with magnitude, shifts of three sigma were common over a plate. Table 3.4 lists sample plates with the maximum shift in the stellar locus between images in the highest 50% of the plate sky-background and the lowest 50% of the sky-background.

At this stage it would have been possible to construct complex corrections to the image parameters as a function of sky-intensity, object magnitude and area for instance. The amplitude, and exact

Table 3.4

Plate	Date measured	Locus shift	Threshold
J757	4/11/79	4.2	9%
J1759C	15/12/79	3.0	10%
J2391C	13/6/80	2.3	10%
J2599	25/10/80	2.9	9%
J2633	8/2/81	2.4	9%
J2652	17/2/80	1.6	8%
J3001	11/1/81	0.7	10%
J3406	25/2/81	1.8	10%
J4445	24/2/81	1.2	7%
J5586	13/5/82	0.8	15%

(a) the errors in the determination of the locus shift are less than 0.2 sigma in all cases. The values refer to the peak change in the shift of the stellar locus over the range $-6 <$ cosmos-magnitude < 0.0, between images in the upper 50% and lower 50% of the plate sky-intensity. All measures were made at 8 micron increment

(b) Plates J1759C and J2391C were found by Corwin (1981) to show defocussing in localised areas. Detailed examination of the parameter shifts with position on the plates clearly show these regions.

dependence on cosmos-magnitude varied from plate to plate, and it was not obvious how to determine the correction reliably for galaxies. The techniques used to quantify the shifts are also insensitive to correlated parameter variations that induce a component of shift parallel to the stellar locus. Changes in image density with sky-background indicated these were also present. Having established that large changes in image parameters were evident in COSMOS measures of type III emulsions as a function of COSMOS sky-background intensity, a more detailed understanding was sought through examination of the measurement procedure.

3.6 COSMOS Calibration and Stability

The calibration of machine measures of photographic plates may be divided into two broad areas. Firstly the determination of the relation between photographic density and the incident light intensity, and secondly the determination of the relation between the machine magnitude measures and photoelectric photometry of stars and or galaxies. In principle if the density to incident intensity relation is well determined, and the properties of the plate and measuring procedure are precisely known - e.g. wavelength response, point spread function, saturation limit and limiting isophote - only a zero point is required and the second phase is unnecessary. In practice this is not the case and a reasonable approximation to the density-incident intensity relation is found with subsequent calibration against photoelectric standards - e.g. Reid and Gilmore (1982). An absolute calibration using photoelectric standards of the UKSTU/COSMOS data was not vital for this project but the achievement of a relative calibration invariant over the whole measured plate area was essential - uniformity is far more important than the absolute scale.

3.6.1 COSMOS Calibration Procedures

To set up the calibration between COSMOS transmission values and incident intensity the following procedure is adopted for each plate. A zero point is adjusted so that the faintest area of sky on the plate corresponds to a T value of about T=230 on the measurement gain - normally gain 1. A region of plate fog is measured, usually at a higher gain due to the low dynamic range at any single gain - this determines the transmission value for "clear plate" - Tc. Step wedge or sensitometer spots - when present - are measured to determine a T value

for each step, wedges are masked to prevent light leakage from the spot halo through the surrounding clear plate. A COSMOS analogue of Baker density (CBD) is then plotted against the relative intensities for the step wedge;

$$\text{Log(Intensity)} = \text{Gamma} * [\text{Log}((Tc-Tb)/(T-Tb) - 1)] + C \qquad (3.1)$$

CBD is the term in square brackets, Tb is the transmission registered by COSMOS when an opaque plate is present in the system. This allows for the detection of light from sources other than directly through the plate, any electronic bias in the system resulting in a current when no light is incident on the detector, or a certain threshold light level being necessary before a detectable current - and hence non-zero T value - is registered. This constant was set equal to zero initially. The constant C is not relevant as an absolute calibration is not attempted. The typically five to twelve points measurable in the CBD vs. Log(I) space are used to determine Gamma: this is accomplished by identifying the region of Log(I) corresponding to the sky level and eye fitting a straight line to the nearby points. The resulting linear relation is used over the entire CBD vs. Log(I) range - i.e. a linear relation between CBD and Log(I) is assumed.

Problems arose because of the difficulty of measuring the calibration steps resulting in incorrect T values, and occasional misdetermination of gamma from the data points. At one time the step wedges were insufficiently well masked and anomalous calibrations severely effected the data. A number of measures had to be discarded after identifcation of the problems but the causes were well understood and remeasures were obtained.

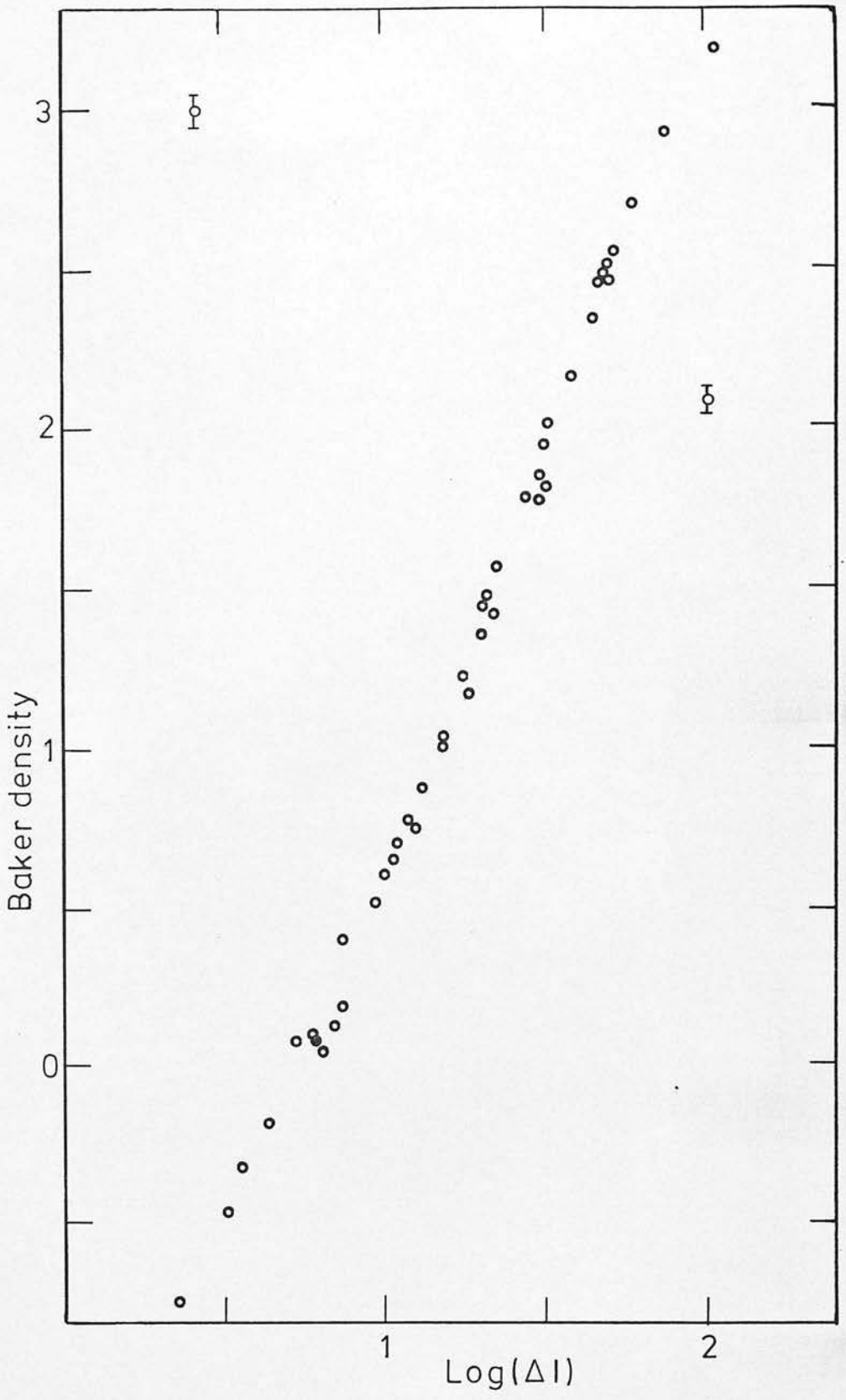


Fig. 3.4 A composite calibration curve for UKSTU IIIaJ plates; data from plates J757, J2652, J3001, J3406 and J4445, measured with a Macbeth densitometer.

3.6.2 COSMOS Measures of Calibration Wedges

Fig. 3.4 shows a composite calibration curve derived from five UKSTU IIIaJ plates. Densities were measured using a two millimeter aperture on the Macbeth densitometer. The Baker density formula (Db=Log(10**(D-Dc) -1) has been used where Dc is the density of chemical fog. This formula is used to linearize the standard Density vs. Log(Intensity) curve near the toe by allowing for the density of unexposed plate. The formula certainly results in an improvement, and for type II emulsions a wide range of the resulting calibration curve is well approximated by a linear relation (de Vaucouleurs 1958). It is evident from Fig. 3.4 that this is not the case for IIIaJ emulsions, a conclusion supported by the data of Dawe (1981) - the COSMOS I.A.M. calibration procedure is not strictly correct.

More importantly no information on how the COSMOS CBD parameter related to density or Baker density measured on a standard densitometer had been obtained. Fifteen plates with COSMOS calibrations were measured using a Macbeth to determine the relation. Several features were apparent; (a) a linear relation between CBD and Db was evident for most plates at intermediate Db, but the slope was considerably different from unity, (b) some wildly discrepant points were evident, particularly at low Db - not due to errors in the Db measurements, (c) evidence for non-linearity was present at high Db, and several curves showed "glitches" in the overall relation. The small number of steps per plate and within the COSMOS density range precluded further analysis, to increase the quality and quantity of data available a request to obtain COSMOS measures of a Joyce Loebl wedge (JLW) and two ANSI standard stepwedges was formulated.

The three wedges were mounted on a clear glass plate masked with opaque paper except for three well separated, one centimeter wide strips corresponding to the wedge positions. This prevented light contributions from the spot halo passing through the glass plate. The JLW (J261) density range was $0.7<D<4.0$ with a gradient of $0.2Dcm^{-1}$. The ANSI wedges consisted of eighteen density steps approximately one square centimeter in extent; a IIIaJ emulsion wedge - $0.2<D<2.41$ - and a IIaO emulsion wedge - $0.2<D<2.1$. In practice mounting restrictions and machine saturation at low and high densities (data was limited to the range $1\leq T\leq 254$), confined the density ranges to $0.8<D<2.4$ (JLW), $0.3<D<2.46$ (IIIaJ) and $0.24<D<1.84$ (IIaO).

Lanes - 128 pixels wide - of data using a 32 micron spot with a 32 micron increment of the IIIaJ wedge at gains 1, 2, 3, and 4, the IIaO wedge at gains 1 and 2, and twentyfour scans of the JLW at gain 1 (one per hour, henceforth designated scans 1-24) were obtained. There were some problems due to large shifts in the COSMOS lane coordinate registration, but these were corrected empirically before further analysis. Data were averaged across lanes to give mean T values (Tm) versus distance along each wedge for each scan. ANSI standard densities were available for the IIIaJ and IIaO wedges, and all wedges were measured using the Macbeth at two millimeter intervals with a two millimeter aperture.

3.6.3 Results

The measures described above were ideal for a variety of investigations of COSMOS performance; signal to noise, transmission drift and consistency of the gain ratios. As these factors are important in determining actual measuring performance they are described before

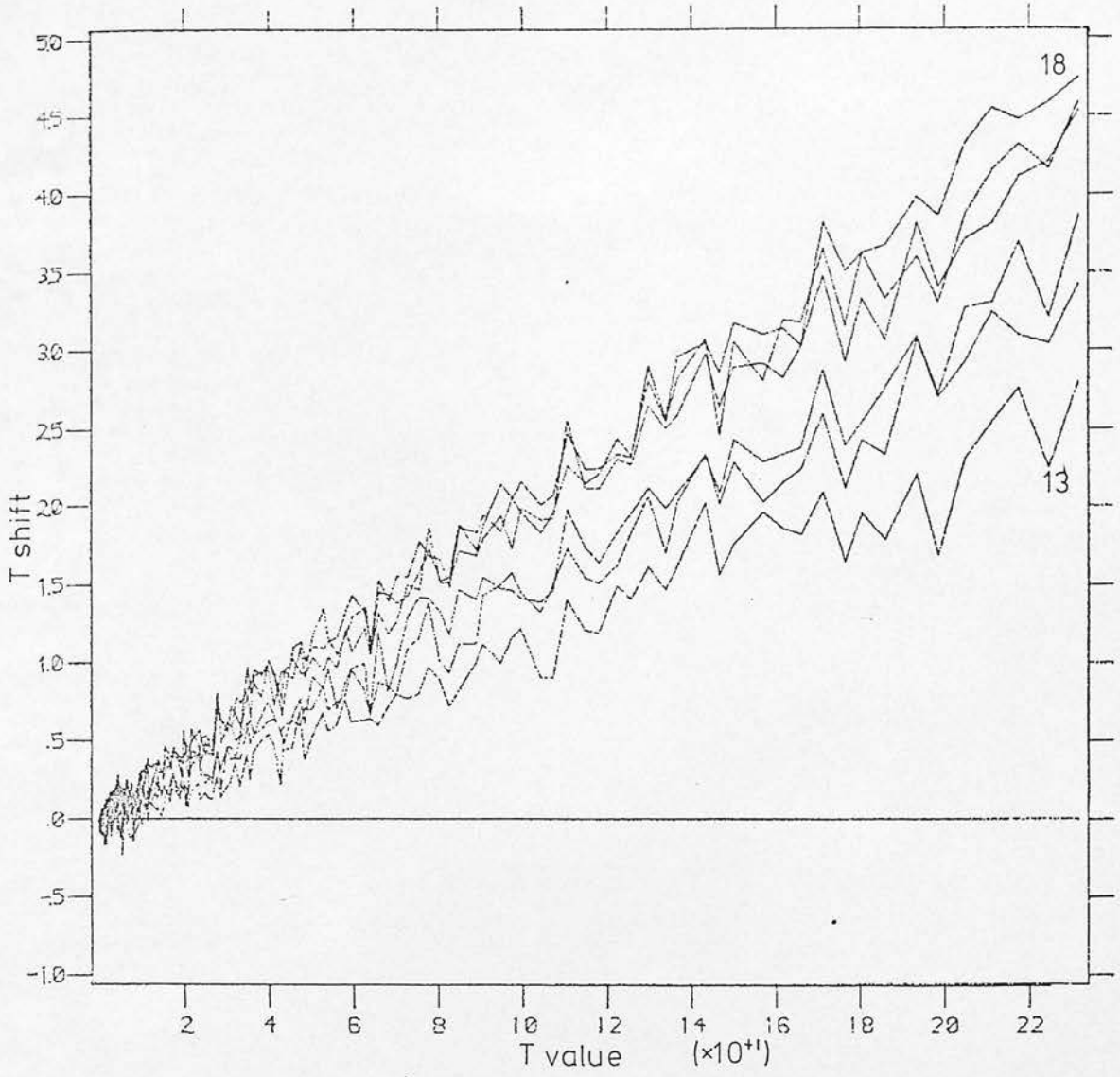


Fig. 3.5 Transmission drift as a function of transmission for six COSMOS scans of the Joyce Loebl wedge (scans 13-18), made at one hour intervals. The curves run sequentially from 13 to 18 from bottom to top of the figure. The measurements were made at gain 1 and are normalised to the result of scan 2.

discussion of the calibration results.

Transmission Drift; The Transmission signal to noise (S/N) at each Y increment was too low for direct comparison between the JLW scans. The data were block smoothed over twenty Y increments to give an average T value (Tav), and a time-series of plots constructed by plotting Tav(scan 2)-Tav(scan x) vs. Tav(scan 2). Scan 2 was chosen, as scan 1 was somewhat anomalous. An example period is plotted in Fig. 3.5 - showing a clear drift of Tav with time - the drift is directly proportional to Tav over the whole Tav range. The rate of drift was not constant with time, showing drift rates of $\leq$0.3% to $\geq$1% over five hour periods. This magnitude of drift was confirmed subsequently by measurement of plate B3499 on successive days with the COSMOS calibration parameters unchanged, and inspection of repeat scans of small areas of plate performed after I.A.M. measures. The conclusion from these sources are consistent with the wedge results, indicating variable drift from 0-2% over a single measure. Providing the calibration relation is linear the effects of this type of drift are removed by the thresholding relative to sky. The rate of drift is compatible with possible variations in the voltages applied to the electronic systems. A record of the room temperature was kept during the measures but no correlation with drift rate was evident. Further investigation would involve monitering of internal cabinet temperatures and input voltages - instrumentation for this purpose was not available.

Signal to Noise; an estimate of the machine signal to noise was made by fitting second order polynomials to the Tm vs. Y scans of the JLW. Fits were made over ten units in Tm - the number of points varying between 80 and 250 due to the non-linear Tm vs. Y relation. The rms scatter from least squares fits was used to derive a S/N estimate. A typical S/N vs.

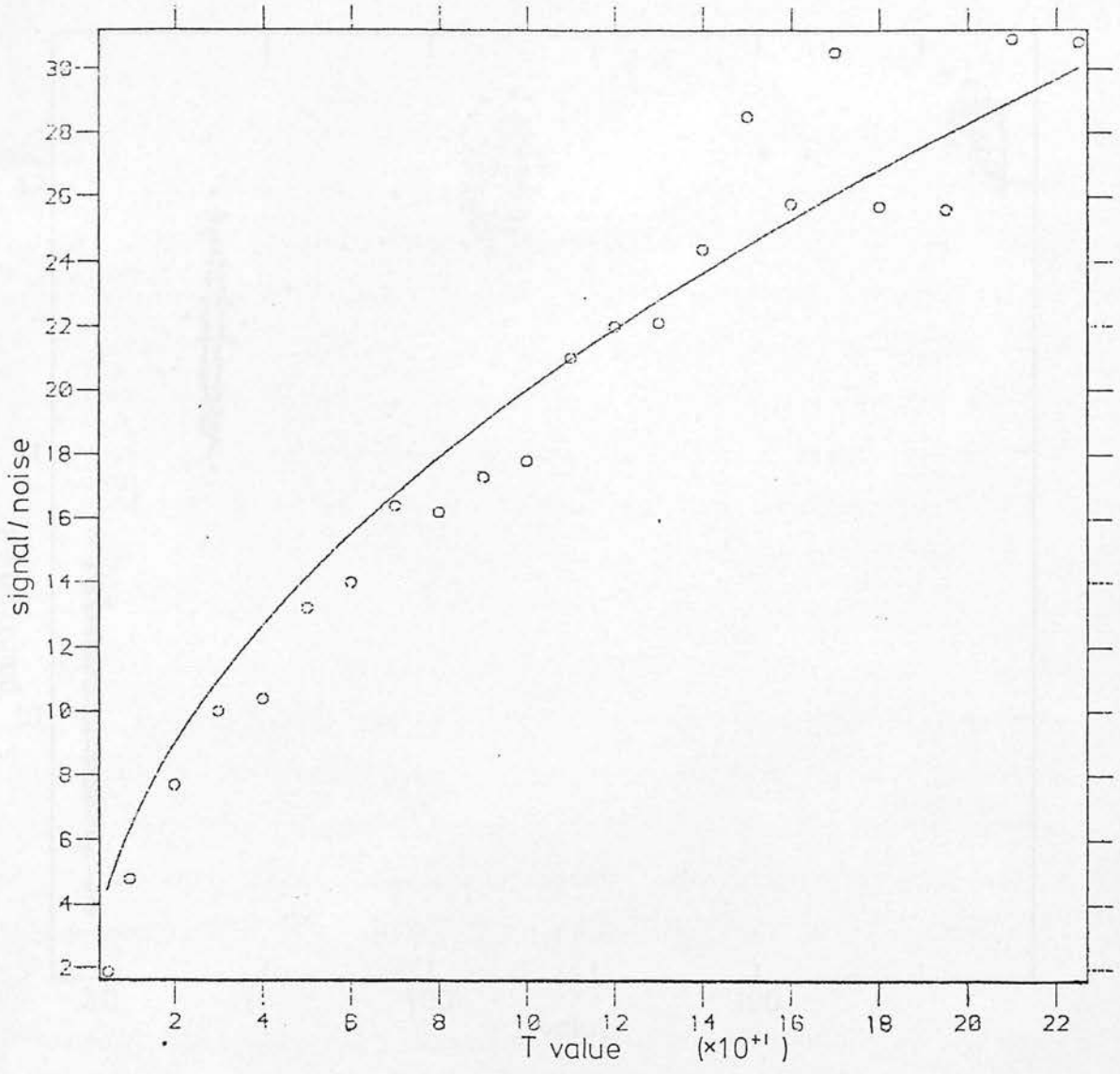


Fig. 3.6 The COSMOS Signal to Noise determination from scan 2 of the Joyce Loebl wedge - the contribution from the wedge has been ignored. The solid line represents the theoretical relation assuming photon noise, arbitrarily normalised at T=240.

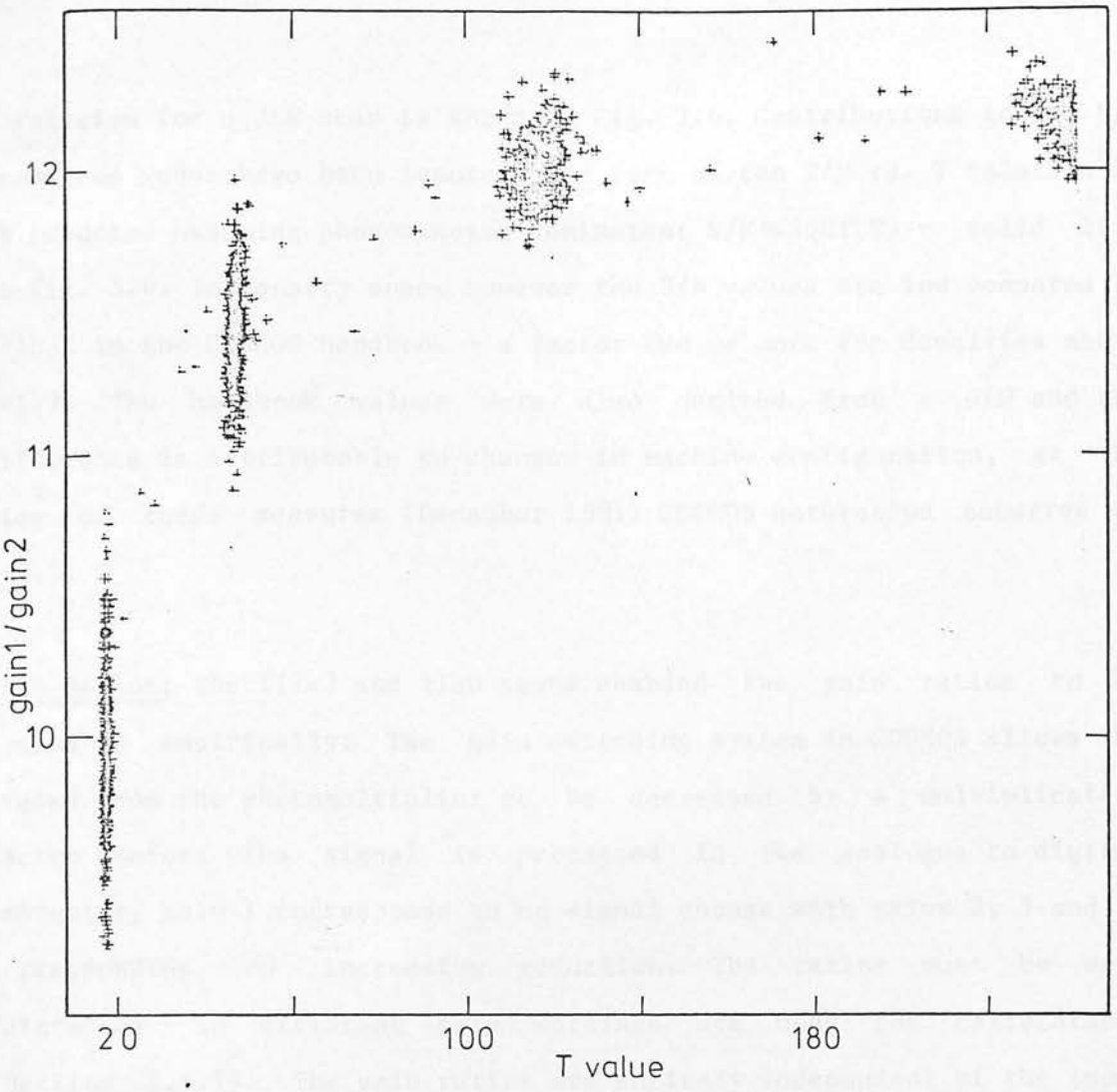


Fig. 3.7 The ratio of the COSMOS gains (gain 1/gain 2) as a function of transmission value (gain 1) from the scan of the IIa0 calibration wedge. The clumping of the data in T space is due to the discrete nature of the density steps.

T relation for a JLW scan is shown in Fig. 3.6. Contributions to the S/N from the wedge have been ignored. The form of the S/N vs. T relation is as expected assuming photon noise dominates; S/N $\propto$ SQRT(T) - solid line in Fig. 3.6. In density space however the S/N values are low compared to values in the COSMOS handbook - a factor two or more for densities above D=1.7. The handbook values were also derived from a JLW and the difference is attributable to changes in machine configuration, at the time of these measures (December 1981) COSMOS saturation occurred at $D \leq 2.5$.

Gain Ratios; the IIIaJ and IIaO scans enabled the gain ratios to be evaluated empirically. The gain switching system in COSMOS allows the signal from the photomultiplier to be decreased by a multiplicative factor before the signal is processed in the analogue to digital converter, gain 1 corresponds to no signal change with gains 2, 3 and 4 corresponding to increasing reduction. The ratios must be well determined - as different gain settings are used for calibrations (Section 3.6.1). The gain ratios are entirely independent of the input signal and should be independent of T value.

The discrete nature of the density steps on the calibration wedges means points are strongly clumped in T space; Fig. 3.7 shows the result for the IIaO wedge in the form gain 1 vs. gain 1/gain 2. At low Tm significant departures from a constant ratio are present - a steep turndown for gain 1/gain 2 (20%), an upturn for gain 2/gain 3 (12%), and a much weaker upturn for gain 3/gain 4 (4%).

The mean values of the gain 2/gain 3 and gain 3/gain 4 ratios are in excellent agreement with those determined from the electronics specification. Gain 1/gain 2 ratio is of the right order (cf. 12.61 from

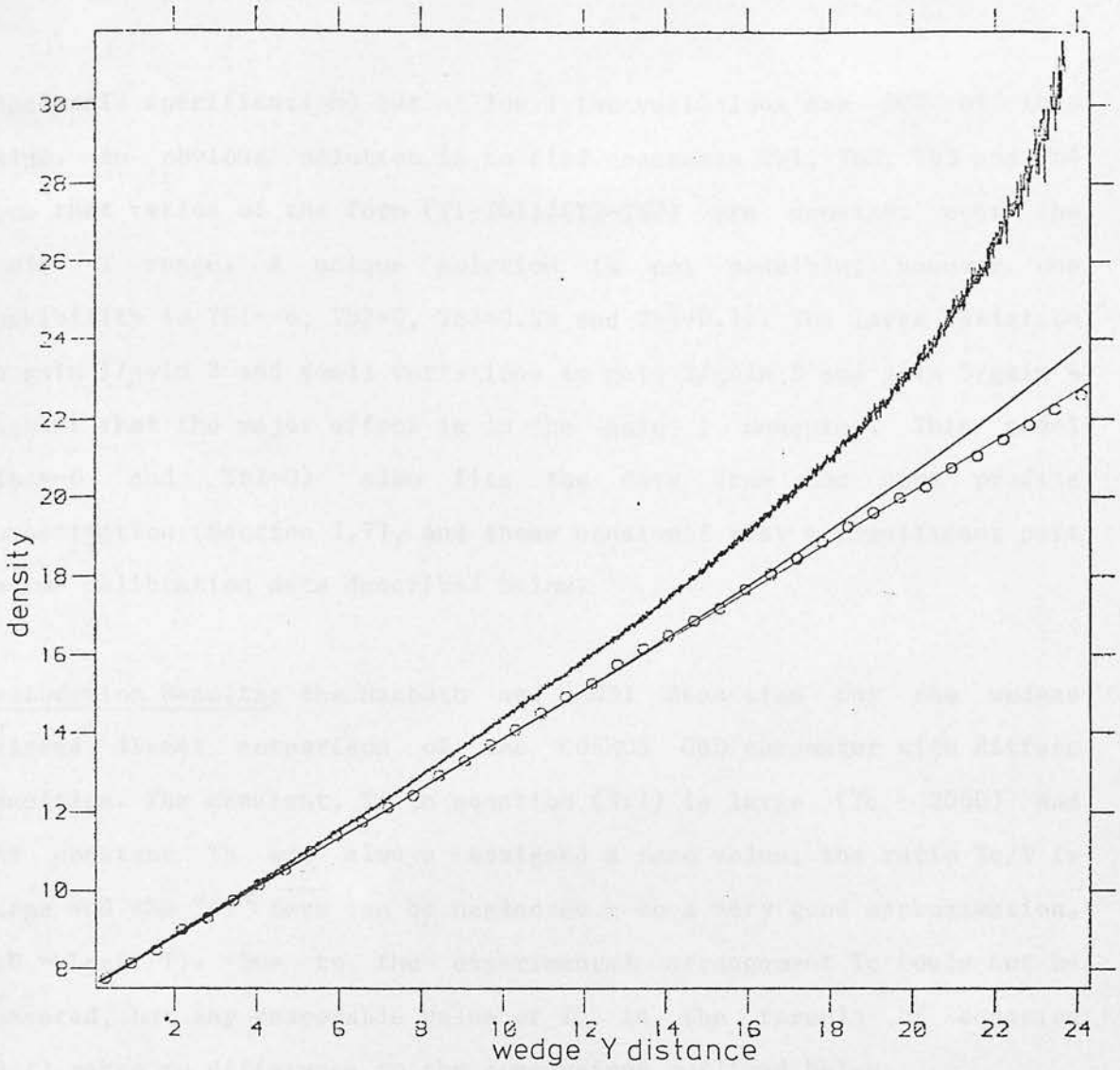


Fig. 3.8 COSMOS Log(1/Transmission) as a function of Diffuse density for scan 2 of the Joyce Loebl wedge - upper solid line. Open circles are the Macbeth measures of the wedge, the lower line, the manufacturer's specified relation. The middle line is the relation for the COSMOS quantity Log(1/(T-Tb)) with Tb =-5. Both the COSMOS relations have been normalised at density 0.8.

electronic specification) but at low T the variations are 20% of this value. An obvious solution is to find constants Tb1, Tb2, Tb3 and Tb4 such that ratios of the form (T1-Tb1)/(T2-Tb2) are constant over the whole T range. A unique solution is not possible, however one possibility is Tb1=-6, Tb2=0, Tb3=0.23 and Tb4=0.12. The large variation in gain 1/gain 2 and small variations in gain 2/gain 3 and gain 3/gain 4 suggest that the major effect is in the gain 1 measures. This model (Tb1=-6 and Tb2=0) also fits the data from the spot profile investigation (Section 3.7), and these constants play a significant part in the calibration data described below.

Calibration Results; the Macbeth and ANSI densities for the wedges allowed direct comparison of the COSMOS CBD parameter with diffuse densities. The constant, Tc in equation (3.1) is large (Tc ~ 2000) and the constant Tb was always assigned a zero value; the ratio Tc/T is large and the "-1" term can be neglected - to a very good approximation, CBD $\propto$ Log(1/T). Due to the experimental arrangement Tc could not be measured, but any reasonable value of Tc in the formula of equation (3.1) makes no difference to the conclusions outlined below.

The linear nature of the JLW means a Density vs. Distance plot should be a straight line - the open circles in Fig. 3.8 are the Macbeth measures of the JLW. The lower solid line indicates the theoretical result assuming the wedge specification is correct. Tests on several JLWs show that J261 shows no significant departures from specification at the level of accuracy considered here (D. Emerson, private communication). The upper curved solid line is the COSMOS density (Log(1/T)) arbritarily scaled to agree with the Macbeth density at D=0.8 - a strong non-linearity is evident.

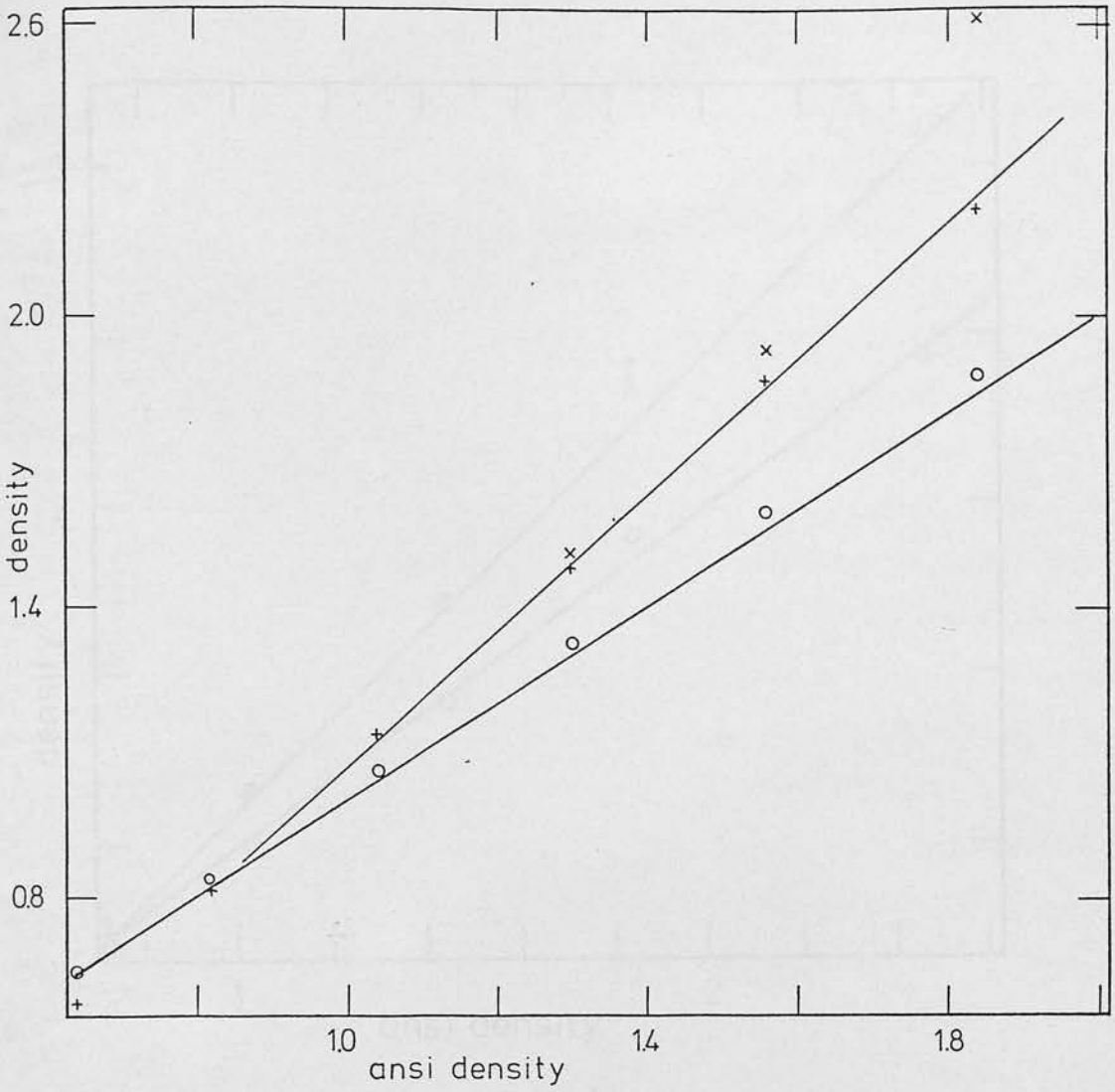


Fig. 3.9 COSMOS values of Log(1/Transmission) as a function of Diffuse density for the IIaO calibration wedge; gain 1 (×), gain 2 (+). The open circles are the Macbeth measures, the lower solid line represents the wedge specification, the upper solid line has slope 1.4 and is drawn for illustration. The COSMOS measures have been normalised at density 0.8

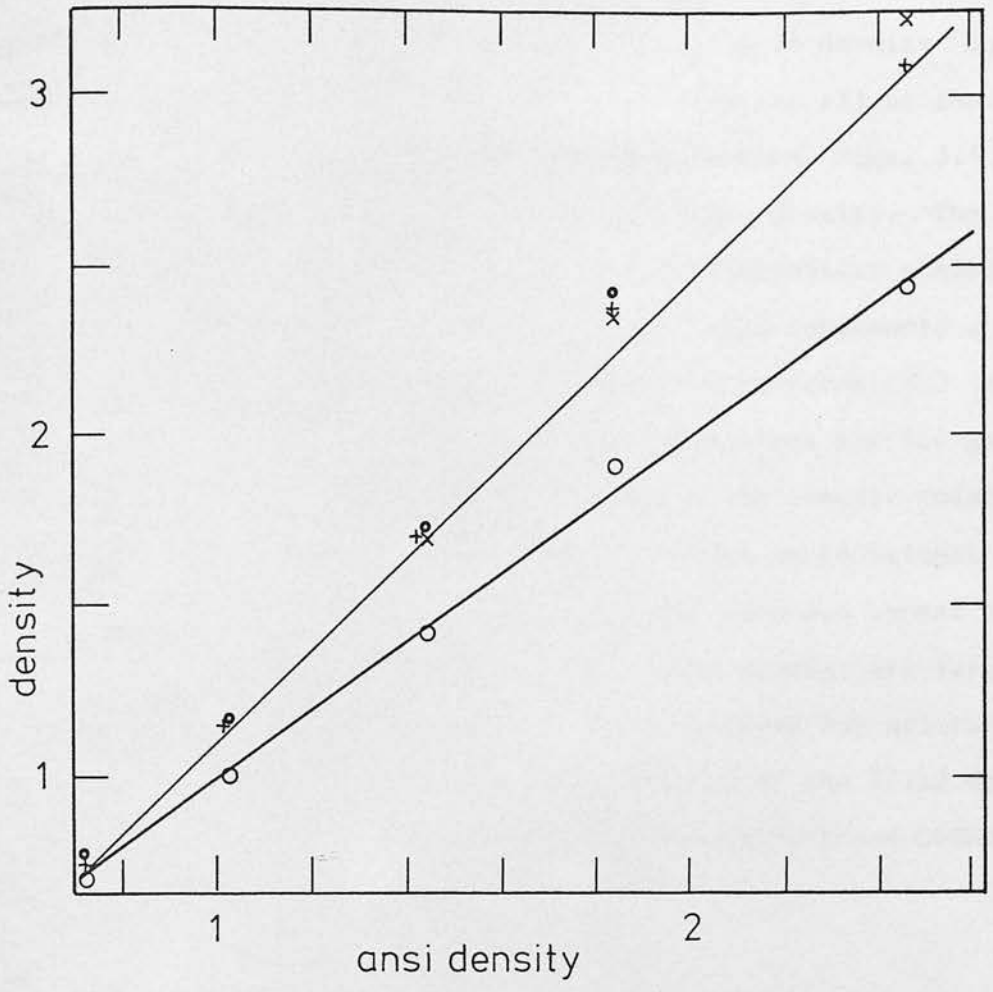


Fig. 3.10 As Fig. 3.9 but for the IIIaJ wedge; gain 1 (×), gain 2 (+), gain 3 (∘). Lower solid line represents wedge specification, open circles are Macbeth measures and upper solid line has slope 1.4.

The ANSI wedges allow only a few determinations of density due to the step quantisation, for some gain settings not all of the steps were registered by COSMOS, restricting the data further. Figs. 3.9 and 3.10 show ANSI density vs. Macbeth and COSMOS density. The Y axis is calibrated in Macbeth density with the COSMOS densities scaled to agree with the Macbeth at D=0.8. The solid line represents a one to one correspondence with ANSI density. The Macbeth measures (o) agree well with the ANSI densities. Note the linear deviations for the gain 2 and 3 measures from the diffuse density relation - the steeper solid line has a slope of 1.4. The gain 1 measures show an additional non-linear deviation in both wedge measures; the effect does not appear as great as for the JLW because the T values for a given density are larger for the ANSI wedges - in transmission space the deviations for all three wedges are entirely consistent. The ANSI density of the IIIaJ density step given as 1.81D appears to be in error as the Macbeth and COSMOS measures on all gains are too high.

In summary the COSMOS density parameter CBD at gain 1 shows evidence for non-linearity in all three wedges. This non-linearity is the same for all measures in transmission space. The CBD measures of the ANSI wedges at all gains show a large linear deviation from the diffuse density relation, the slope of the CBD vs. Density relation being approximately 1.3-1.4.

3.6.4 Discussion

These results confirm the earlier indications from the plate stepwedge data that the COSMOS calibration embraces several features not previously recognised. The transmission drift detected is considerably larger than the nominal specification but within the range of variations

expected from electronic considerations. The explanation of the lower S/N results and early onset of saturation is almost certainly due to machine configuration changes since the earlier determination. The linear deviations of COSMOS CBD from diffuse density are explicable in terms of measurement of semi-specular rather than diffuse density, though no specific discussion of this was available before the measurements. The cone angle the photomultiplier subtends at the emulsion is ~10 degrees, data in the Kodak "Plates and Films" handbook and from similar experiments performed on Japanese microdensitometers (Okamura and Davenhall 1982, IDPU internal report) are available: these show the values of Specular/Diffuse density for IIaO and IIIaJ emulsions in the density range examined are as expected - COSMOS measures semi-specular density with a cone angle of 10 degrees. This is consistent with the steep CBD vs. Log(I) and CBD vs. Db slopes observed in data from plate stepwedges. No component of specular density is evident in the COSMOS JLW measures because a JLW is effectively a neutral density filter and absorbs rather than scatters light.

Two explanations for the non-linearity evident in the gain ratios, and in the gain 1 measures of the wedges were suggested: (a) bit dropping within COSMOS, (b) a zero point offset within the machine. Bit dropping arises because COSMOS determines transmission values to ten bit precision within the analogue to digital converter, but the final output is to eight bit precision - the least significant bits being dropped. In the absence of noise the eight bit output will form a "staircase" above the ten bit output in the CBD vs. diffuse density plots. In practice the presence of noise means the eight bit output is systematically biased to lower T values - the effect is most significant at low T. The magnitude of the effect is determined by where the zero point for the output T values is set relative to the ten bit data in the analogue to digital

converter. If the zero point is set as T=1 (output) corresponding to T=4 (ten bit analogue to digital output) the effect is insignificant. Software simulation of the worst possible case (i.e. T=1 output set to T=1 ten bit output), converting the eight bit data to six bit data showed that bit dropping was not responsible for the major portion of the observed non-linearity.

The non-linearity could be explained by a large negative Tbl (i.e. Tb in equation 3.1), physically this would correspond to a certain threshold intensity of light being necessary before a value of T=1 is recorded. The nominal COSMOS set up procedure is such that the system oscillates on the border of T=0/T=1 for no light present in the system. If due to a negative Tbl, the non-linearity would be strongly dependent on T value, independent of density and may vary with time and gain setting - all of which fit the observations. Fig. 3.8 shows the COSMOS measure of the JLW reduced with a Tbl=-5 in the CBD formula, a Tb value of this magnitude will empirically remove the non-linearity. However attempts to detect the offset within the COSMOS electronics suggest that the offset magnitude is confined to $-2 \leq Tbl \leq 0$ contradicting the empirical result. COSMOS time to procede beyond this point was unavailable and this disagreement was not resolved.

The COSMOS calibration procedure was altered so that a Tb=-6 was adopted as the default value for all I.A.M. measures from December 1981. The empirical results were incorporated in a simulation of the interaction between the COSMOS measuring and calibration procedures with variable background plates - described in section 3.8.

3.7 Determination of the COSMOS Spot Profile

No detailed study of the COSMOS spot had been succesfully undertaken within the five years to June 1982, the determination of the spot profiles is essential to understand the interaction of COSMOS with a photographic plate exhibiting a large dynamic range in density. The COSMOS machine allows the use of three spot sizes - nominally 8, 16 and 32 microns full width at half height (FWHH). Signal to noise investigations of photographic emulsions (e.g. Okamura et al. 1983) and the ability of the machine to measure high density pixels are particularly sensitive to the spot core profile shape, and halo extent. The spot was checked crudely every week by eye examination of a razor edge profile, but the data was only of sufficient quality to detect gross anomalies in spot parameters. A previous attempt to obtain the spot profile was made by McNally (1979) using a deconvolution technique, this was not succesful due to the low S/N of the data. To obviate this difficulty a different approach was taken - modelling the data utilising simple assumptions concerning the spot profile. The availability of a 7 micron slit, and a razor edge enabled differential and integral line spread profiles to be obtained for the three spots. The measures described below were made in May 1982.

The 7 micron slit was oriented almost parallel to the COSMOS Y-axis, and the spot repeatedly scanned across the slit giving some fifteen hundred scans consisting of 128 T values at 8 micron intervals almost parallel to the X-axis and normal to the slit. The slit was rotated until parallel to the X-axis and the spot scanned in successive lanes over the slit giving lanes of data with 128 T values parallel to the slit. The razor edge was scanned in an identical fashion to the scans of the slit parallel to the X-axis. Scans were performed at gains 1 and 2 to

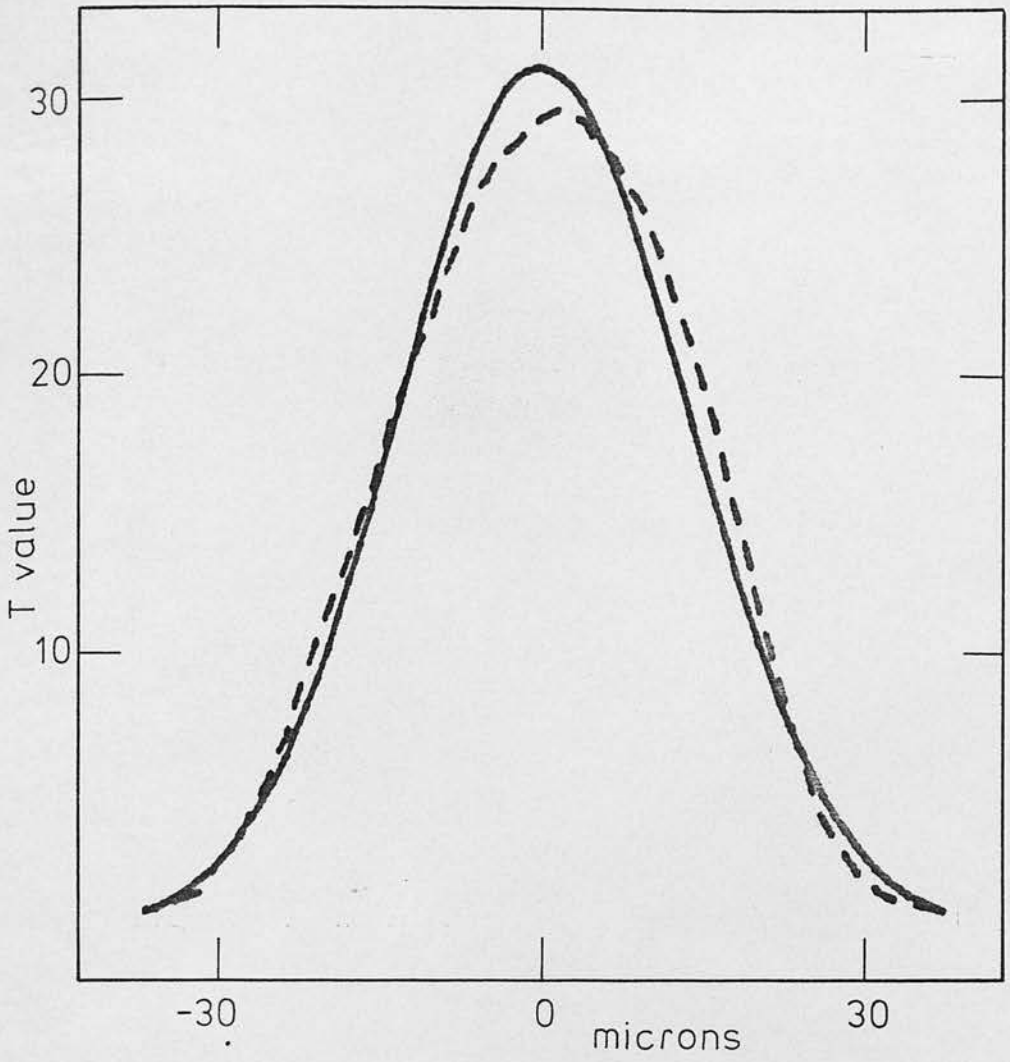


Fig. 3.11 The COSMOS line spread profile for scans of the 32 micron spot across a 7 micron slit - measures made at gain 2. The solid line is the line spread profile for the adopted gaussian spot core model.

increase the transmission resolution of the data.

3.7.1 Data Reduction

The Spot Core; for reasons explained below the COSMOS scans with the slit parallel to the Y-axis were most suitable for the spot core determination. The small residual slope of the slit relative to the Y-axis was determined from the XY positions of the peak T values in each scan. For each spot/gain combination the distance from the slit center was determined allowing for the slopes, and mean T values from a thousand scans gave a slit profile at 1 micron resolution. Distance bins for scans at gain 1 containing saturated pixels (T=255) were eliminated, and the data was truncated when more than 2% of the T values in a distance bin contained T=0 pixels. The latter operation prevents the systematic bias towards low T values due to a threshold light intensity being necessary before a T=1 value is recorded. If the gain ratio is constant with T value, a simple multiplicative factor should relate the gain 2 data to the gain 1 data. The results confirmed earlier findings that the ratio is a function of T value - a negative constant Tb1=-6 was required with Tb2=0 to make the ratio (T1-Tb1)/(T2-Tb2) constant over the entire T range. Small changes were evident in the value of the multiplicative constant for 8, 16 and 32 micron spots - 13.4, 12.4 and 12.25 respectively. This effect is similar to the variable factor observed in the ANSI wedge measures and may be due to electronic drift on timescales of order an hour. The gain 1 data provided no improved resolution at low T values due to the high S/N of the Gain 2 data, the remaining analysis is confined to the gain 2 data. The slit profile for the 32 micron spot is shown as the dashed line in Fig. 3.11. The noise is small over the whole T range due to the large number of scans used to generate the profiles.

Gaussian fits to the spot profiles were made using a non-linear least squares fitting procedure (Bevington 1969, p234) to model the slit profile. The calculation of error functions and other numerical integrations were performed using NAG library routines. Circular symmetry was assumed in the modelling process. The photocathode receptor area is large enough - 2.5cm diameter - that diffraction effects at the edges of the slit may be ignored. The slit profile generated from the best fit gaussian for the 32 micron spot is shown as the solid line in Fig. 3.11. The parameters of the fitted gaussians are given in Table 3.5, the fit to the 8 micron spot was not good, but a model incorporating two gaussians gave an excellent fit. The integrals under the gaussian fits to the three spots agree to better than 0.5% indicating the success of the fitting procedure and that the same fraction of light in each of the spots is contained in the core. The core of the 32 micron spot is close to specification but the 8 and 16 micron spots are significantly larger: spot sizes are 13.7, 26 and 32 micron FWHH..Comparison of data obtained from the smaller spots using these profiles is in good agreement with previous S/N estimates for COSMOS and photographic emulsions, (e.g. Okamura et al. 1983) - results assuming the specified spot profiles were highly discrepant.

The slit data from scans parallel to the slit were not as useful: the slit had to be scanned in 128 pixel wide lanes and the lane coordinate registration was not good enough to simply abut lanes to give a large number of scans. Further, the spot scans at a small angle to the X-axis which broadens the slit profile. However, allowing for this broadening, the data is in excellent agreement with the previous slit data, indicating the spot cores are symmetric to better than 10% - consistent with the earlier assumption made concerning symmetry of the spot core.

Table 3.5

Spot core parameters

Spot	FWHH (microns)	Percentage of light	
8	13.9±0.2	95.0±0.5%	Gaussian
8	11.1±0.2	77.0±0.5%	Gaussian
	27.9±0.3	23.0±0.5%	Gaussian
16	25.5±0.3	95.0±0.5%	Gaussian
32	31.1±0.4	95.0±0.5%	Gaussian

Spot halo parameters

uniform disc	radius 4.40±0.05mm	percentage of light	2.95±0.05%
uniform disc	radius 2.05±0.05mm	percentage of light	2.05±0.05%

The small but systematic asymmetry evident in the slit profile of Fig. 3.11 is primarily due to the time constant of the photocathode response which results in a slight T enhancement for the data to the right of the slit profile peak.

The Spot Halo; the slit data was of no use for the halo determination as an integration over 7 microns far from the spot center is insufficient to register a T=1 value. The six scans of the razor edge were averaged at each Y increment to give a mean T vs. distance from spot center relation. The gain 1 data saturated when the spot center approached within 30 microns of the razor edge, contained no additional information, and was discarded.

Beyond 50 microns from the spot center the edge profiles for the three spots were indistinguishable within the noise. Each profile displays the same anomalous features; small decreases in the integrated edge profile at several points. Fig. 3.12 shows the edge profile for the 16 micron spot with several of the features indicated. The explanation of the

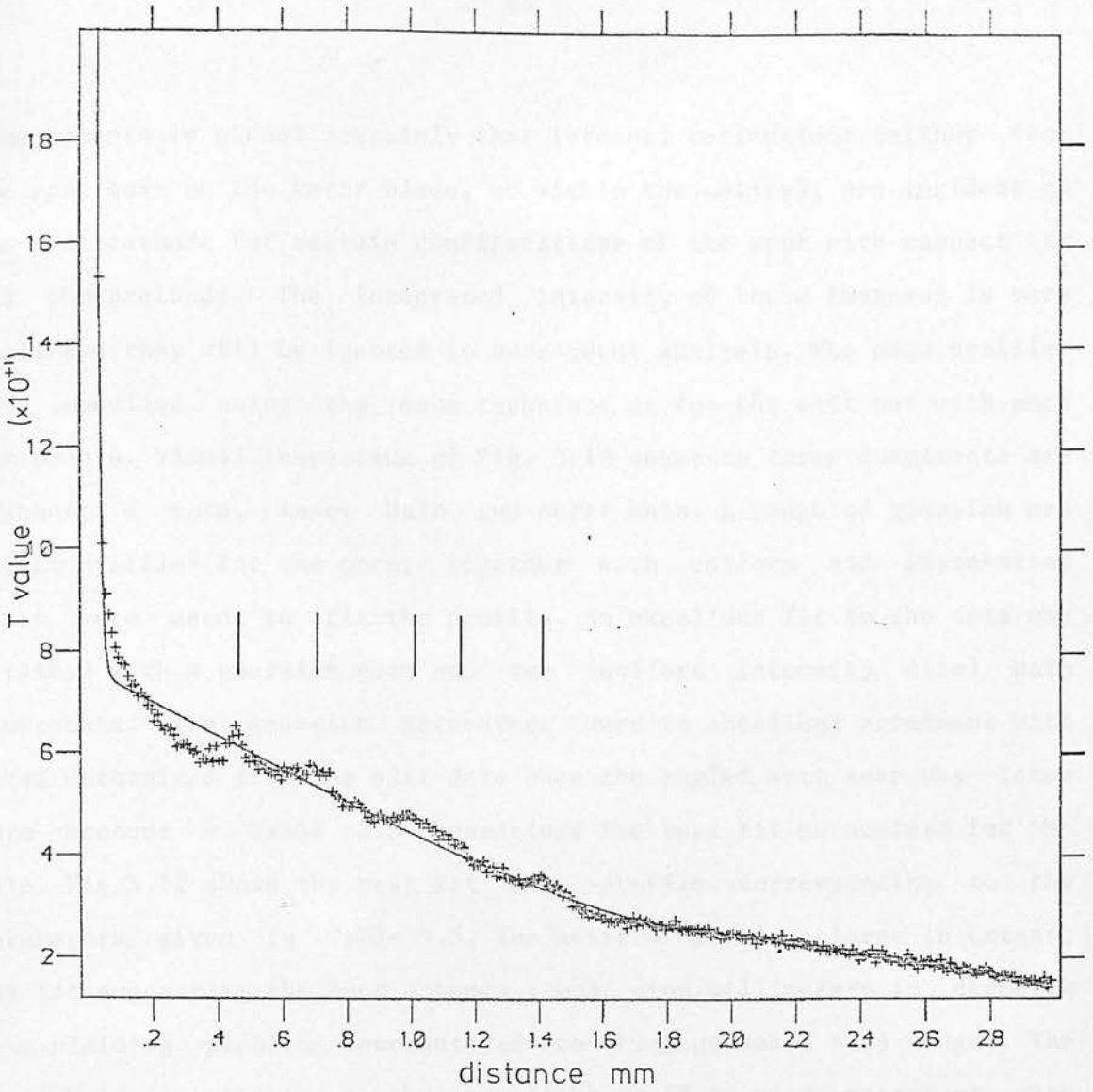


Fig. 3.12 The COSMOS edge profile for the 16 micron scans of a razor edge - measures made at gain 2. Distance from razor edge vs. integrated transmission is plotted. The smooth solid line is the edge profile of the adopted core-halo model for the spot - consisting of a gaussian core and two uniform intensity discs. The vertical bars indicate the positions of anomalous features in the edge profile probably due to internal reflections.

enhancements is almost certainly that internal reflections (either from the spot core on the razor blade, or within the optics), are incident on the photocathode for certain configurations of the spot with respect to the photocathode. The integrated intensity of these features is very small and they will be ignored in subsequent analysis. The edge profiles were modelled using the same technique as for the slit but with more components. Visual inspection of Fig. 3.12 suggests three components are present; a core, inner halo and outer halo. A range of gaussian and Moffat profiles for the core, together with uniform and exponential discs were used to fit the profile. An excellent fit to the data was obtained with a gaussian core and two (uniform intensity disc) halo components. The gaussian parameters were in excellent agreement with those determined from the slit data once the angled spot scan was taken into account - Table 3.5 summarises the best fit parameters for the halo. Fig 3.12 shows the best fit edge profile corresponding to the parameters given in Table 3.5. The uniform discs are large in extent, the fit suggesting the spot extends almost nine millimeters in diameter - explaining problems encountered measuring unmasked step wedges. The total halo contribution to the spot light is 5% in good agreement with the observation that the effective saturation limit for very large images measured by COSMOS occurs at T ~ 12 in typical I.A.M. scans. Saturation is expected at T~ 0.05*Tsky for an object blocking the core light - Tsky is set to ~ 230 in I.A.M measures.

The fit to the halo is certainly not unique, though the two uniform discs were much the best fits of any model tried. The close agreement between the core gaussians from the slit and edge profiles is encouraging and the halo fit is extremely good as shown in Fig. 3.12. If an improved fit was to be made, the halo components would show a decrease in intensity near the core to model the very flat portion of

the profile between 100-400 microns.

3.8 Simulation of COSMOS Measurement Procedure

A computer simulation to examine the consequences of the non-linearity and spot size data obtained in the previous sections was developed. The aim was to examine the interaction between the COSMOS measurement procedure and images present on varying sky-backgrounds, together with effects of possible background misdetermination.

Moffat and gaussian image profiles in intensity space were used throughout; little was gained by convolving model galaxy profiles with point spread functions, and gaussian profiles with larger dispersions were used to represent galaxies. Images with associated sky-backgrounds were generated in intensity space, the profiles were converted to density space using a mean calibration relation for IIIaJ emulsions - Fig. 3.4. A relation between CBD and diffuse density (Section 3.6) was used to convert profiles to CBD. The spot halo was incorporated in the simulation to model the density saturation limit of COSMOS. Values of summed intensity and cosmos-magnitude were then obtained assuming a linear CBD vs. Log(I) relation and applying a threshold cut to the profiles. This procedure incorporated the main factors operating in the COSMOS measurement of photographic plates. In addition the effects of the non-linearity at low density of the IIIaJ Baker density vs. Log(I) relation and possible misdeterminations of sky-background by COSMOS were examined.

Fig. 3.13 shows the positional change of images in the cosmos-magnitude vs. Log(area) parameter space caused by a 3% under-determination of the sky-background, confirming that errors in sky-background determination

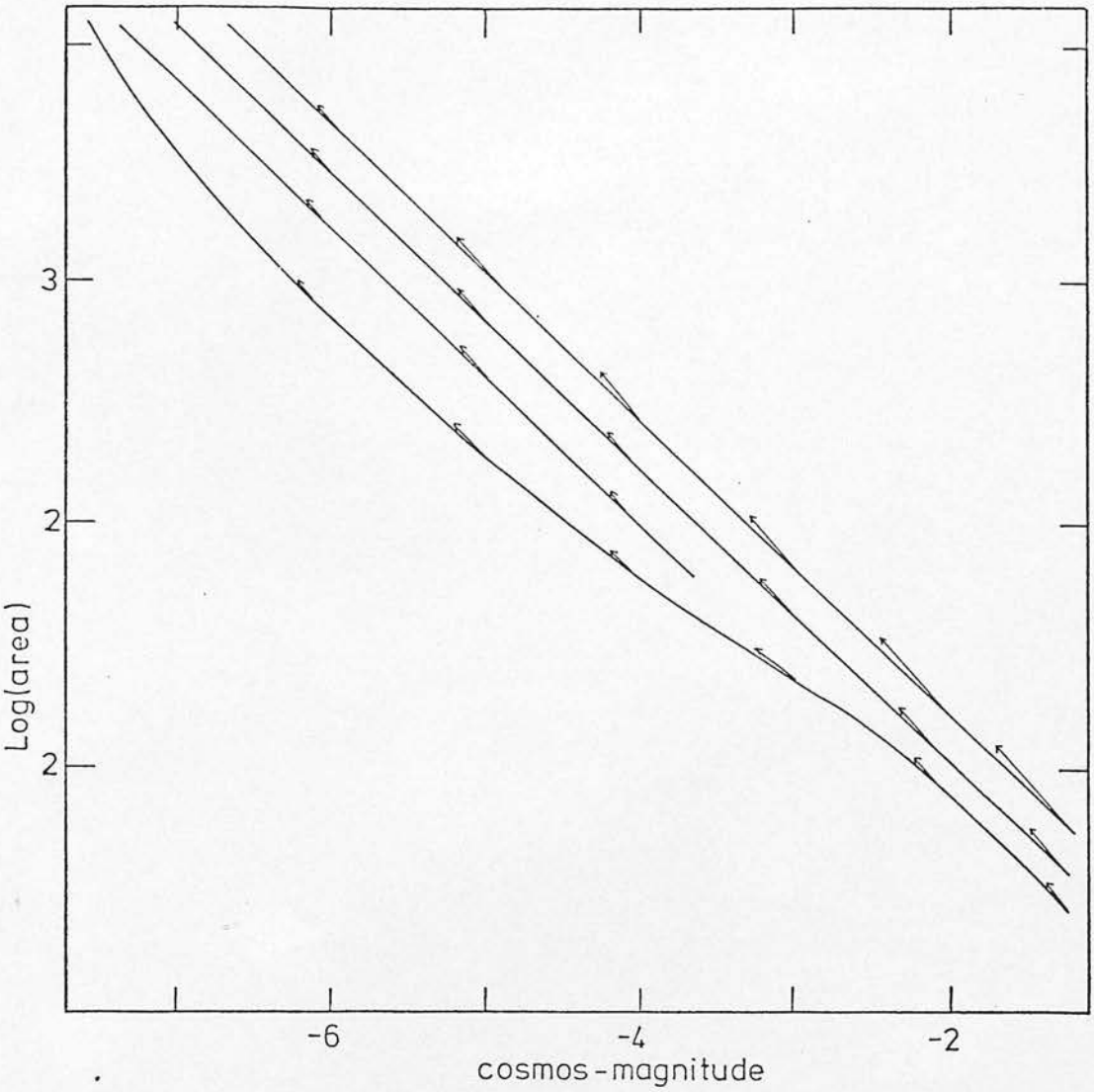


Fig. 3.13 The positional changes for gaussian images in cosmos-magnitude vs. Log(area) parameter space. The vectors illustrate the magnitude (curves represent the initial image parameters) of the changes due to a misdetermination of the sky-background by -3%. The lower curved line represents the stellar locus, and the straight lines above, the region of parameter space occupied by galaxies.

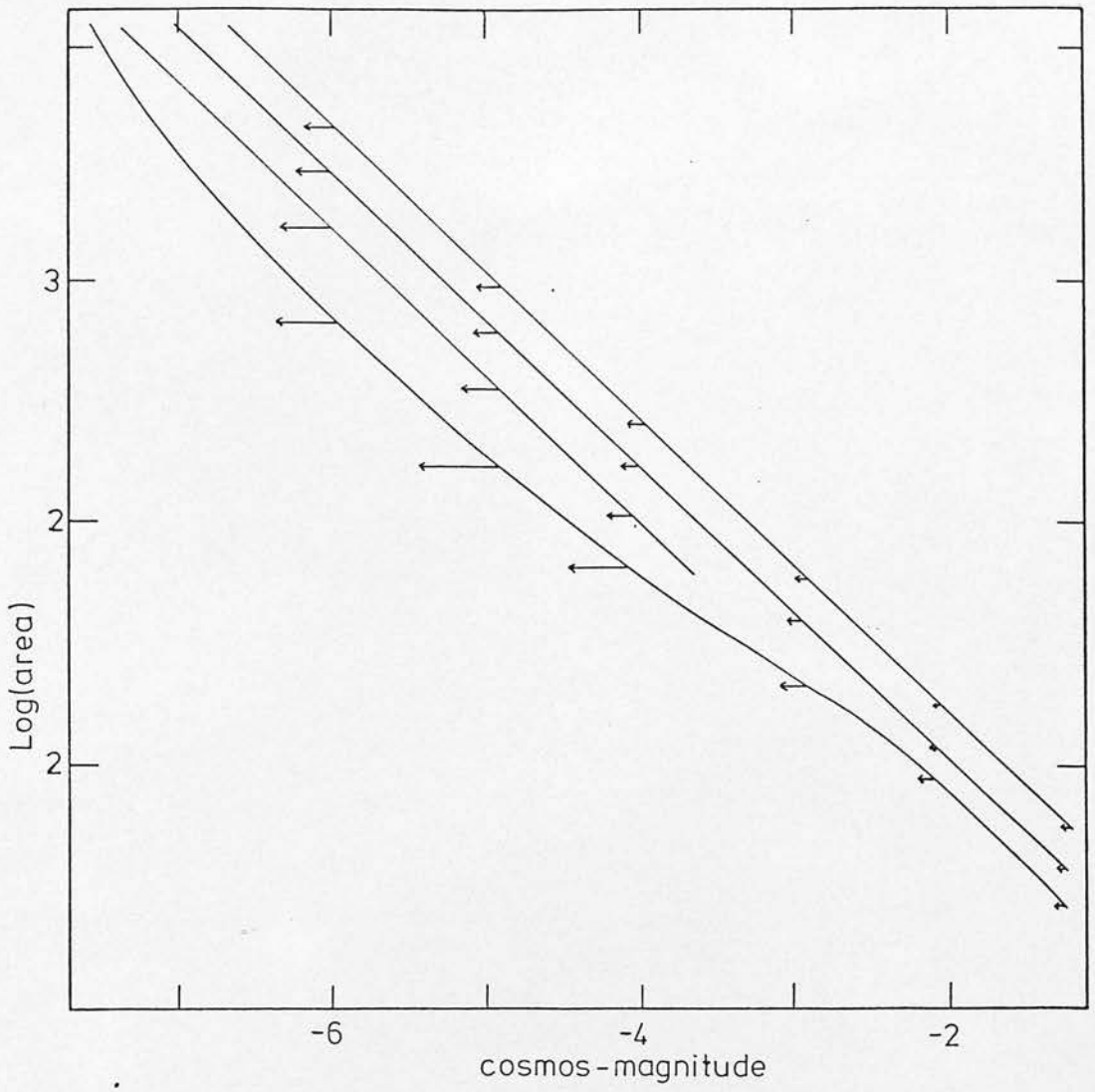


Fig. 3.14 The positional changes for gaussian images in cosmos-magnitude vs. Log(area) parameter space. The vectors illustrate the magnitude (curves represent the initial image parameters) of the changes due to a non-linear calibration in the presence of a 5% sensitivity change.

produce correlated changes in image magnitude and area parameters, a conclusion that was verified in other parameter spaces. This rules out the possibility that all parameter shifts in the COSMOS data were due to problems with the background following routines. Fig. 3.14 shows the changes evident for a 5% sky-background change with the non-linearity at high density incorporated. The shifts agree well with those observed in COSMOS data - compare Fig. 3.14 with Fig. 3.3 - the non-linearity modifies the profiles of bright saturated images and the effect is strongest for stars. The deviation reaches a maximum at some cosmos-magnitude then decreases for brighter images; COSMOS transmission saturation becomes important (images become fainter at high sky-intensity), and the onset of detectable diffraction spikes adds a lsb component to the brighter images - this lsb component is largly unaffected by the calibration non-linearity. The exact form of the field effects for bright images is very sensitive to the COSMOS saturation limit - a factor that can vary considerably from measure to measure. These effects explain the characteristic shape of the field effects on the stellar locus illustrated in Fig. 3.3.

In practice both calibration non-uniformity and sky-background misdetermination are present in the data. The excellent agreement between the form of the parameter shifts due to non-linearity at high density and the change in number density of images caused by sky-background misdetermination explains the predominant systematic parameter shifts evident in COSMOS measures up till late 1982. The illustration above employs one particular non-linear calibration relation, the extreme sensitivity of field effects to the calibration means that a whole range of field effects can be generated by "reasonable" changes in the calibration relation - if Tb varies between ± 6 this could explain the amplitude and form of most field effects seen

in COSMOS measures.

3.9 COSMOS Specification

From the beginning of 1982 the COSMOS I.A.M. output was sufficiently stable that a specification for the COSMOS machine was desirable. This was necessary to determine the accuracy of the data used in the clustering investigation - a useful by-product was the first reliable determination of COSMOS performance over large areas of plate. The data used here is taken from pairs of COSMOS measures listed in Table 3.6. More detailed results from the pairing test programs have been made available to IDPU, with much of the data incorporated in the 1982 COSMOS handbook.

Table 3.6

Plate	Measures	Increment	Threshold
V2499	26/5/82	8	10%
	14/5/82	16	10%
	15/5/82	16	10%
J3593C	8/12/81	8	7%
	6/6/82	8	10%
J5586	9/5/82	8	15%
	21/7/82	16	15%
J6124C	7/12/81	8	10%
	12/6/82	8	10%
R6230	24/3/81	8	15%
	10/2/82	16	10%
	11/2/82	16	10%
UJ7149C	10/5/82	8	10%
	11/5/82	8	10%
UJ7154C	21/2/82	8	10%
	24/2/82	8	10%

The specification relates to I.A.M. data obtained at a 10% threshold cut from IIIaJ and IIaD emulsions. The data represent averages over all image types - results for stellar images are significantly better (e.g. Reid and Gilmore 1982), and those for galaxies somewhat worse. Small changes in the threshold level between the measures, combined with limitations of the pairing software result in approximately 2% of images not appearing in the statistics. These images are nearly all members of very close groupings on the plate. The total number of images detected by COSMOS also varies because of threshold changes - variations of up to 30% in total image density are evident between "identical measures" of the same plate, illustrating the danger of not considering completeness limits carefully. With the current pairing software the COSMOS multiple measures are complete to 96-98% at cosmos-magnitude 0.0 on IIIaJs at a 10% threshold cut, and to cosmos-magnitude -1 on IIaDs at 10% threshold cut. Reliable comparisons between COSMOS measures are not available below these levels.

Table 3.7 gives results for the major positional, shape and photometric parameters. In many cases systematic changes in parameters are evident between pairs of measures and rms estimates are given for data with these global systematic changes removed. All data refers to errors expected on a single measure. The rms of the major parameters as a function of cosmos-magnitude are shown in Fig. 3.15; the data are averages of the repeat measures of plates J6124C and UJ7154C which are typical of high quality IIIaJ plates measured by COSMOS. The horizontal bars that lie above the histograms in Fig. 3.15 represent the values of the rms error after allowance for the fraction of images lost in the pairing procedure. The following general comments can be made concerning the results in Table 3.7:

Table 3.7

positions: systematic shifts over measure ± 6 microns, no magnitude dependence
no asymmetry in amplitude of rms, or systematic errors between the X and Y coordinates
rms 10 4 4 5 microns

sky-background: systematic shifts over measure 0-2% relative intensity
rms 0.3-0.5% relative intensity

cosmos-magnitude: systematic shifts over measure 0.0-0.15, weakly magnitude dependent
systematic shifts between measures 0.0-0.5, strongly magnitude dependent, especially for faint objects
rms 0.2 0.04 0.06 0.2

peak surface brightness: systematic shifts over measure 0.0-0.12, weakly magnitude dependent
systematic shift between measures 0.0-0.4, strongly magnitude dependent
rms 0.12 0.06 0.07 0.11

Log(area): systematic shifts over measure 0.0-0.06
systematic shifts between measures 0.0-0.2, strongly magnitude dependent
rms 0.08 0.02 0.04 0.08

ellipticity (UW): systematic shift between measures <0.01
rms 0.04 0.04 0.06 0.14

ellipticity (IW): systematic shift between measures <0.01
rms 0.03 0.03 0.05 0.12

orientations (UW): systematic shift between measures <0.4 degrees

rms	2.0	4.5	6.5	10.0	$e > 0.5$
rms	3.0	4.0	8.0	13.0	$0.5 < e < 0.3$
rms	5.0	7.0	12.0	15.0	$0.3 < e < 0.1$

orientations (IW): systematic shift between measures <0.4 degrees

rms	1.5	4.0	6.0	9.0	$e > 0.5$
rms	2.0	3.0	7.0	12.0	$0.5 < e < 0.3$
rms	3.5	6.0	11.0	14.0	$0.3 < e < 0.1$

(a) rms values in groups of four are for images with cosmos-magnitudes -6, -4, -2 and 0 in that order.

(b) systematic shifts over measures reflect the stability of the machine over a single measure - changes manifest themselves as "field effects".

(c) systematic shift between measures represent the consistency of COSMOS when attempting "identical" measures of the same plate. Changes in threshold level and calibration changes are the main cause of the strong magnitude dependent shifts between measures.

Table 3.7 (cont)

(d) the ellipticity limits used in the orientation rms calculations are intensity weighted.

(e) "strongly magnitude dependent" is defined as a shift of at least 0.25 magnitudes or 0.1 in Log(area) over the cosmos-magnitude range -6 to 0.

Positions; systematic effects are evident in both X and Y coordinates as a function of plate position. These occur for measures made in the same orientation as well as rotated through ninety degrees, and the systematic errors are independent of object size or magnitude. The errors occur either at the COSMOS pixel position registration stage, or in the offline pairing software. The errors given in Table 3.7 however, are representative of the data a COSMOS user would receive using the standard reduction software. Fig. 3.15d illustrates the amplitude of the positional rms errors as a function of cosmos-magnitude.

Sizes; systematic image size parameter errors are essentially a measure of how precisely the threshold level is defined over and between measures. The rms errors in Log(area) - Fig. 3.15b - are representative for all sizes, and unit weighted best fit ellipse parameters.

Shapes; systematic changes between ellipticity measures are entirely dependent on how accurately the threshold level is defined. The lower the threshold level the more quickly the images will tend to a peaked distribution about intermediate ellipticity with fainter magnitudes. The random errors in shape parameters (Fig. 3.15e) are extremely consistent from measure to measure.

Orientations; the stability of the orientations relative to the COSMOS coordinate system is extremely good, even for cases where plates were

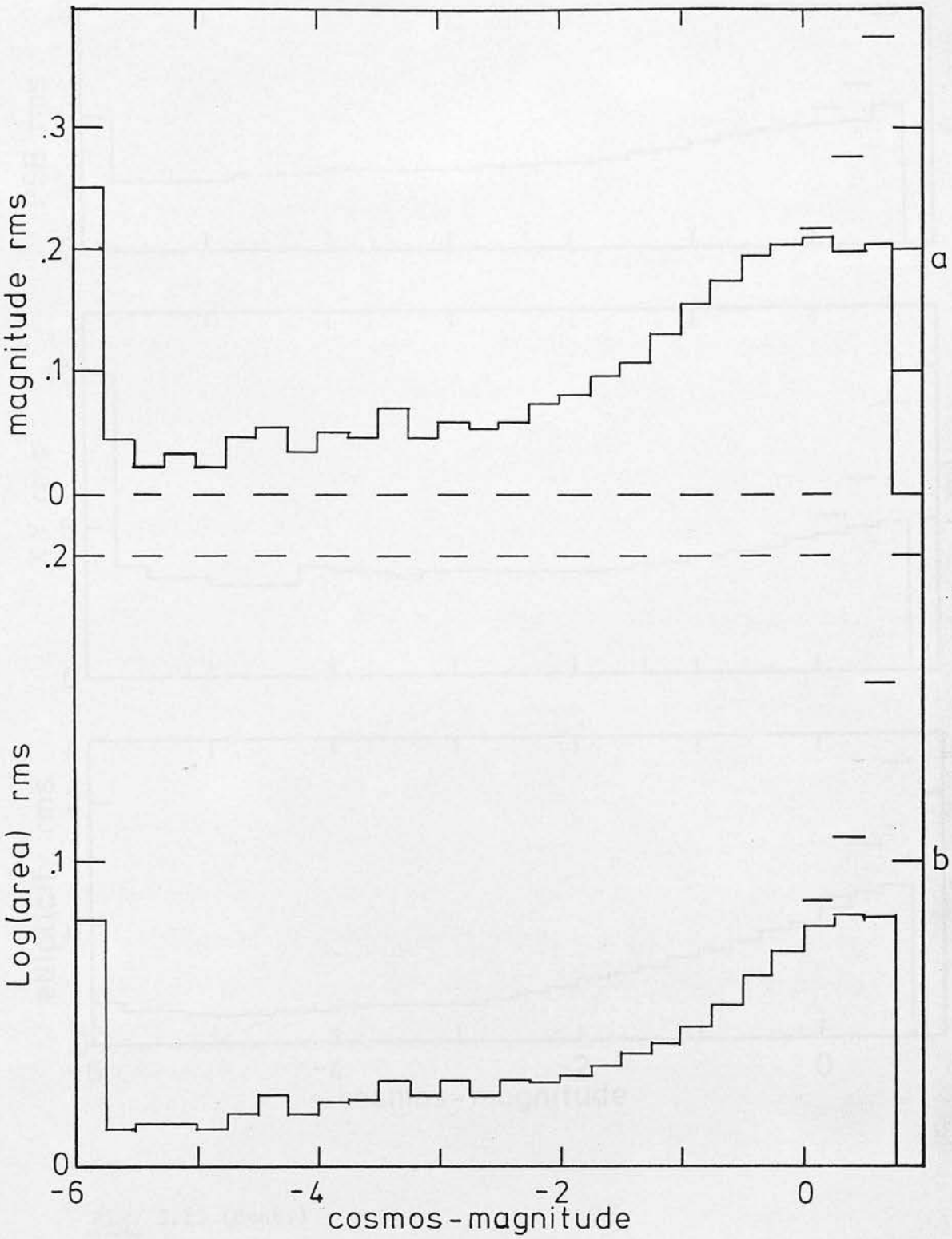


Fig. 3.15 RMS errors for COSMOS I.A.M. parameters as a function of cosmos-magnitude. The data are averages from repeat measures of plates J6124C and UJ7154C. The horizontal bars above the histograms at faint magnitudes represent the calculated rms after allowance for the loss of images in the pairing process.
(a) cosmos-magnitude rms (magnitudes)
(b) Log(area) rms (Log(number pixels))
(c) peak surface brightness rms (magnitudes)
(d) positional rms (microns)
(e) intensity weighted ellipticity rms (ellipticity)

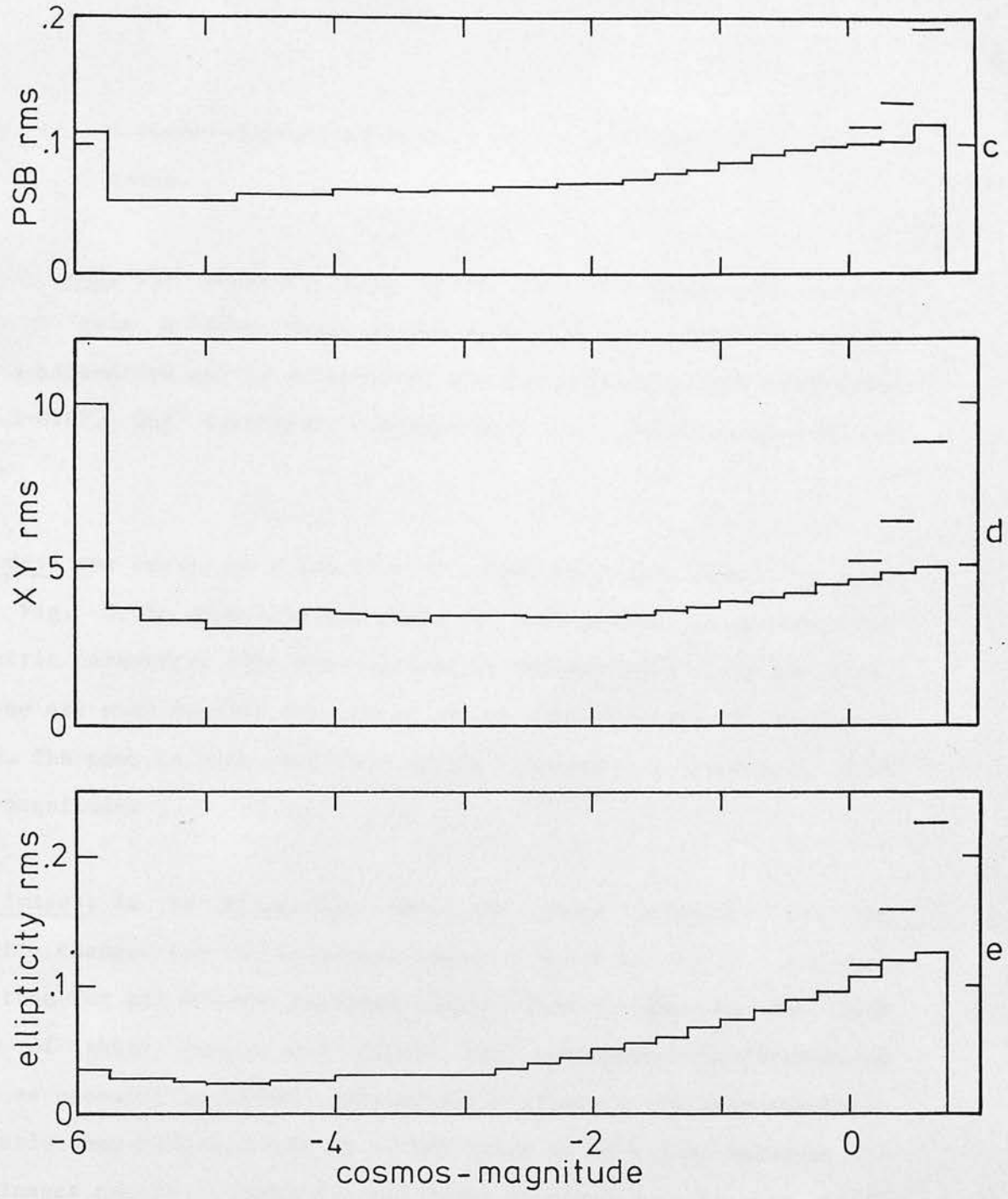

.2
.1
0
PSB rms
c
10
5
0
XY rms
d
.2
.1
0
ellipticity rms
e
-6
-4
-2
0
cosmos-magnitude

Fig. 3.15 (cont.)

rotated through ninety degrees systematic errors are less than half a degree in all cases.

Sky-background; the accuracy with which the sky-background can be determined sets a lower limit to the precision with which photometric and size parameters can be determined. The rms sky-background variations are 0.3-0.6%, but systematic changes with plate position of 1-2% are normal.

Magnitudes; rms errors as a function of magnitude are shown in Fig. 3.15a, Fig. 3.15c shows the rms error for peak surface brightness. The photometric parameters also show systematic changes with plate position, but these are much smaller than those encountered prior to the beginning of 1982. The peak to peak amplitude of the changes is typically 0.15 cosmos-magnitude.

Bright Images; it is noticeable that the image parameter rms and systematic changes for the brightest images - $m_B<16$ and $m_V<14$ - are much larger than for all but the faintest images. This is due to the high density of their constituent pixels and consequent low information content as measured by COSMOS. Photometric parameters are very sensitive to exactly how COSMOS is set up at the start of an I.A.M. measure; the bright images contain a large dynamic range in pixel density and small changes in calibration result in very different intensity profiles. As the lsb components of the images are often asymmetric large positional shifts also occur. The brightest images significantly contaminate the sky-background determinations on occasions, which can cause large variations in parameters between measures - the use of fainter images to accomplish coordinate transformations for plate pairing, and conversion from XY coordinates to right-ascension/declination would result in

significant improvements in accuracy.

3.10 Summary

The data from the COSMOS machine showed dramatic improvements over the period of this project. Some of this improvement was due to changes in the COSMOS configuration; installation of an autofocus mechanism, increase in Transmission resolution and correction of errors in background filtering for example. Large systematic parameter shifts in the data at the start of the project were identified and quantified for the first time. The systematic changes in parameters were found to correlate with the local sky-background present on type III emulsion plates. An examination of the properties of these systematic changes combined with new data on aspects of the COSMOS measuring procedure enabled an explanation consistent with the observations to be found. The new data, relating to the shape and extent of the COSMOS spots, the relation between COSMOS Transmission and diffuse density were particularly important in this regard. A simple computer simulation allowed the systematic parameter shifts in the data to be quantitatively modelled, results showing that non-linearity evident between the COSMOS density approximation and diffuse density was a significant cause of parameter shifts.

With the correction of the non-linearity and improvements in the nature of the COSMOS background following method, the COSMOS I.A.M. data attained some stability. This allowed an empirical determination of the COSMOS machines performance specification for the first time - although field effects are still detectable at a much reduced level in multiple scans of single plates. This investigation shows that the COSMOS machine is potentially capable of producing data from large areas of plates to

sufficient accuracy for galaxy clustering investigations to be attempted.

4 UKSTU Plates

4.1 Introduction

There is no doubt that the quality of UKSTU plates is vastly superior to any other wide field photographic material available. Even intermediate quality plates show large increases in limiting magnitude and uniformity when compared to the POSS glass copies, or current ESO and Palomar plates. However the demands of current quantitative astronomical research are severe and it is necessary to achieve very high degrees of uniformity over areas of 25 square degrees. The importance of this was stressed in the introduction to this work. The study of large scale galaxy clustering is concerned with very low amplitude, large scale effects not considered by earlier workers. Little quantitative data on this type of plate variation is available. After the initial discovery of large field effects in the COSMOS measures of AAT and UKSTU plates an attempt to ascertain the nature of any contribution attributable to the plates was made, checks were also performed on the completeness limits of the UKSTU plates.

In spite of the large quantity of plate material now available from the UKST, detailed investigation of the variation in the quality of astronomical data from the plates is difficult: few fields possess multiple plates with a wide range in quality, and it is difficult to obtain glass copies and large allocations of automated machine time to perform the analysis. Further, there is a lack of superior plate and photoelectric data to establish an "absolute" measure of performance, although preliminary work has been performed as a by-product of galaxy cluster (Carter 1980) and image classification investigations - Chapter 5. The investigation described in this Chapter should be regarded as

highly preliminary, and much remains to be done as more material becomes available and measuring machines become more reliable.

Calibration of UKSTU plates has been investigated in some detail and recently composite calibration curves for different emulsions have been given by Dawe (1981), but no quantitative data on the variation in shape or changes in sensitivity as a function of plate position are available. Distortions in image shape and intensity profiles due to telescopic effects such as field rotation are well documented (Wallace and Tritton 1980). The overall grading of UKSTU plates - the quality control - is carried out to a strict specification relating to a number of factors, but little data on the quality of astronomical data obtained from plates of different grades (i.e. plate quality) was available. For the detection of large scale low amplitude clustering, variations in detection criteria and astronomical object parameters with plate position must be very small. This requires that variations in background fog, additive sky-background intensity and shape of the calibration relation are constrained within tight limits.

4.2 Plate Background Variations

A phenomenon noticed early on was the gross differences evident between the large scale - order of degrees - sky-background variations on plates of different emulsion type: type II emulsions, show no detectable background changes over the central five degree diameter area, at larger radii the rapid drop off in background density due to telescopic vignetting is evident. In contrast direct hypered type III emulsions often exhibit large variations in sky-background over their entire area. The variations show an approximate radial dependence - the maximum density is attained near the plate center. The variations are not

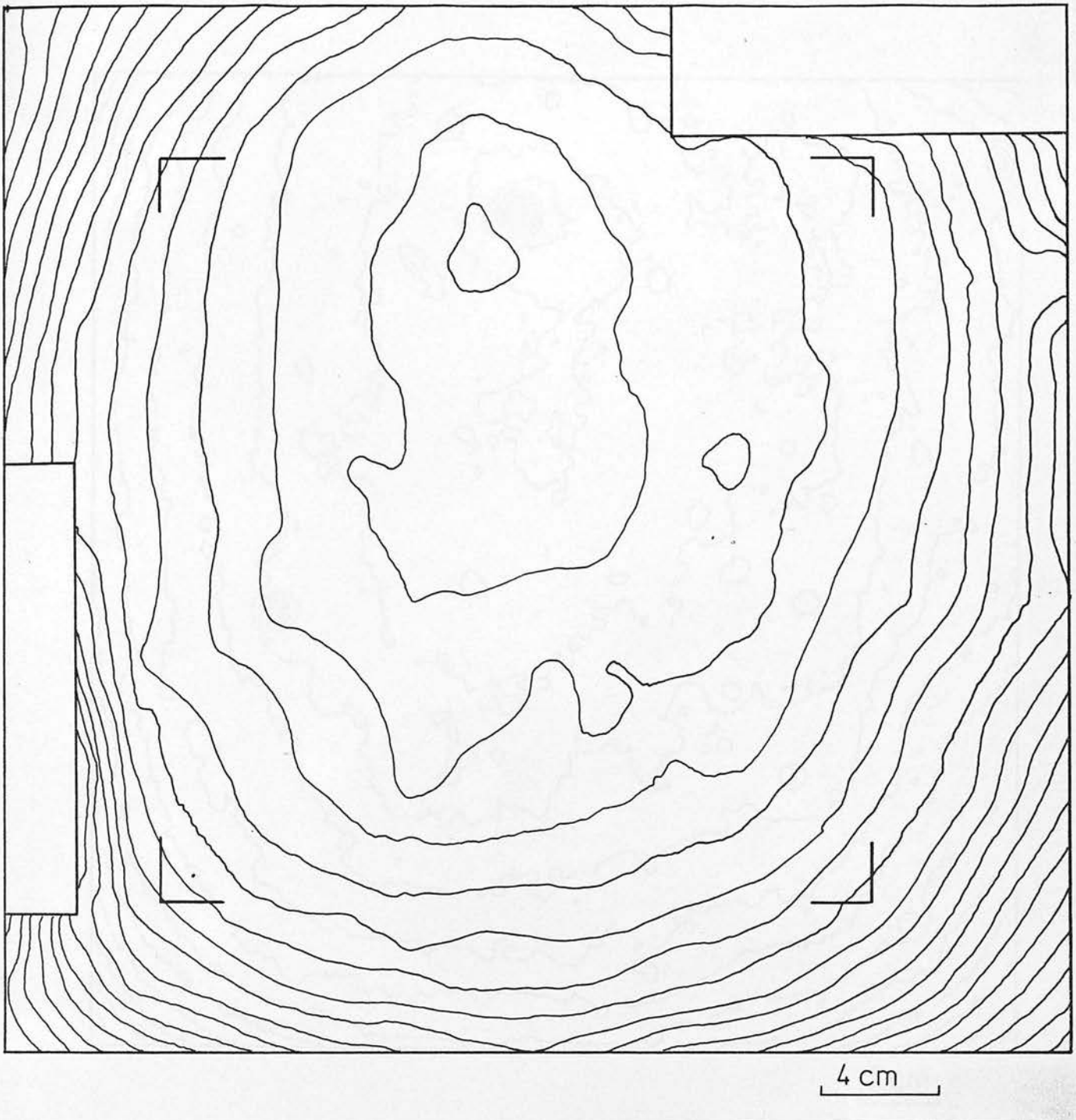


Fig. 4.1a The sky-background for plate J2633 plotted in relative intensity - contour interval 1%. The blanked out areas correspond to the plate stepwedge positions. North is to the top, east to the left.

(b) The sky-background for the area of J2633 measured by COSMOS - contour interval 1% in relative intensity. The extent of the COSMOS area is indicated by the box corners in Fig. 4.1a.

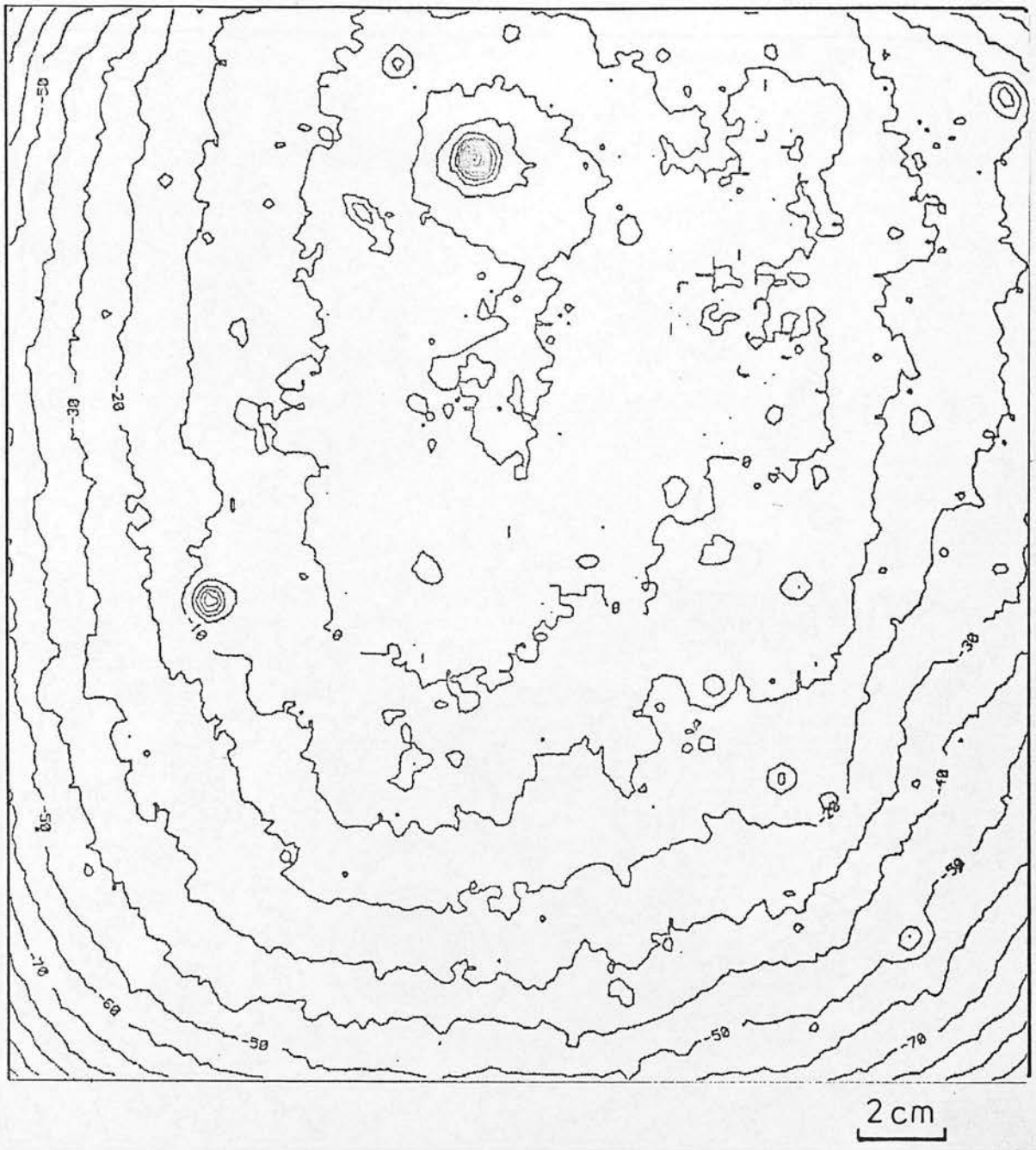


Fig. 4.1b

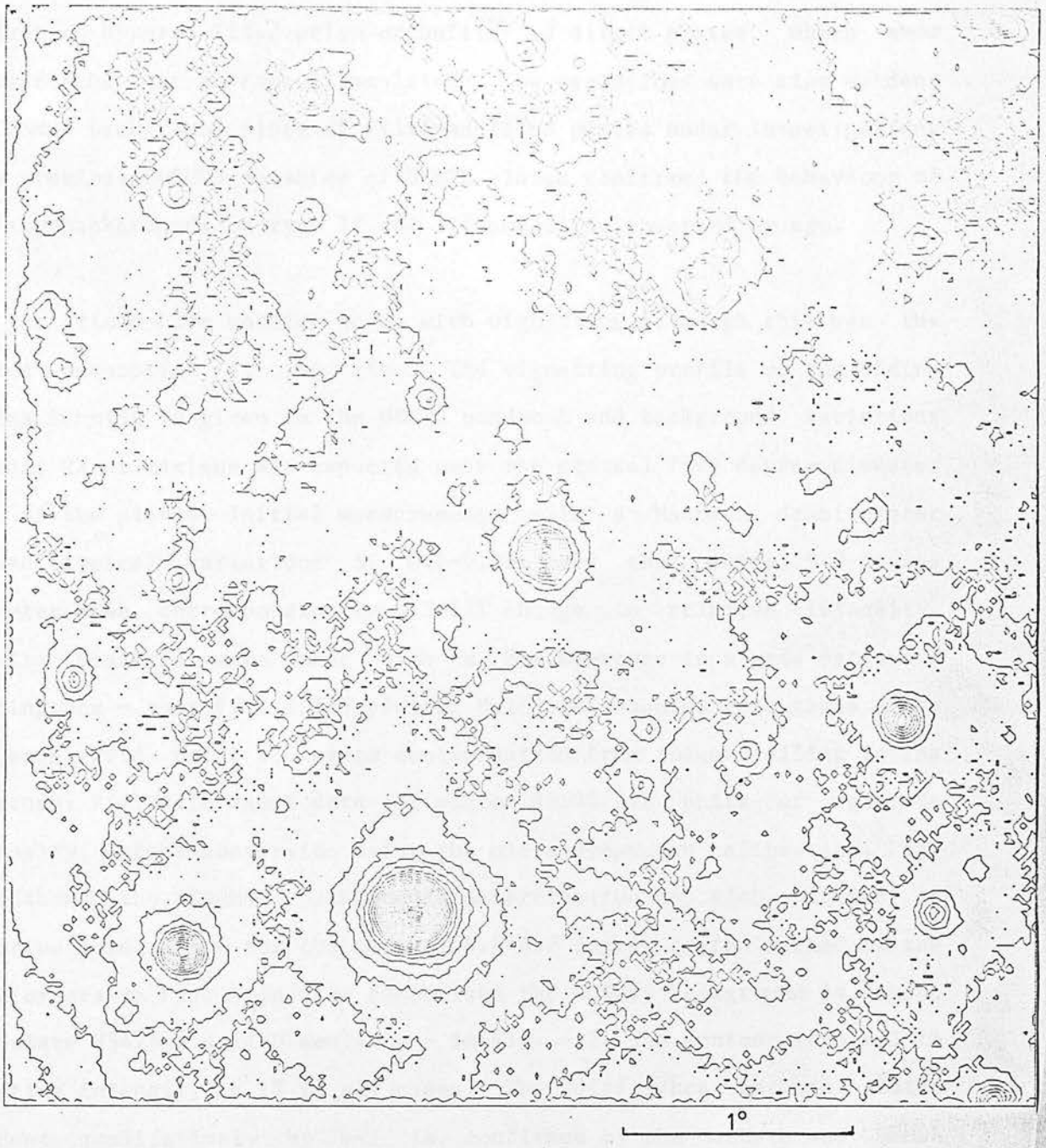


Fig. 4.2 The sky-background for plate V3475 from COSMOS - contour interval 1% in relative intensity. The small regions of dense contours are bright images which modify the sky-background. North is to the top, east to the left.

evident on hypered IIIaJ prism or unfiltered direct plates, which show similar behaviour to type II emulsions. The variations were also evident in COSMOS background plots of filtered IIIaJ plates under investigation. Many previous COSMOS measures of UKSTU plates confirmed the behaviour of the sky-backgrounds on type II and III emulsions observed by eye.

The variations have nothing to do with vignetting although this was the common assumption at the time. The vignetting profile of the Siding Spring Schmidt is given in the UKSTU handbook and background variations of only 2% at maximum are expected over the central five degree diameter area of the plates. Initial measurements with a Macbeth densitometer showed typical variations of 0.1-0.2D over the central 5.0 degree diameter area, corresponding to a 5-12% change in relative intensity. Density measures were made with a 2mm aperture in a grid pattern - spacing 2cm - over five IIIaJ plates. Multiple measures were taken close to each grid point to reduce contamination from images falling in the aperture. Fig. 4.1a shows data for plate J2633 in units of relative intensity, after conversion using the plate stepwedge calibration. Fig. 4.1b shows the COSMOS background determination - also plotted in relative intensity - for the central 4.6*4.8 degree region marked by the box corners in Fig. 4.1a. For comparison the COSMOS background is shown for plate V3475 - a IIaD emulsion - in Fig. 4.2. The contour interval in relative intensity is 1% in all cases. The difference between plates evident qualitatively by eye is confirmed by the Macbeth and COSMOS measures. Table 4.1 lists the variation in background intensity within a radius of 2.5 degrees of the plate center - excluding regions contaminated by bright stars - for five UKSTU J plates.

Table 4.1

Plate	Density change	Intensity change	Fog change
J757	0.12±0.02	6±1%	0.03±0.01
J2633	0.15±0.02	7±1%	0.03±0.01
J3001	0.08±0.02	5±1%	0.03±0.01
J3406	0.12±0.02	6±1%	0.02±0.01
J4445	0.15±0.02	7±1%	0.02±0.01

Further conversations with members of the UKSTU revealed that the effect on IIIaJs had been noted by D. Malin while preparing high contrast prints of UKSTU plates. Malin's explanation was in terms of the differential loss of emulsion sensitivity caused by the varying amount of water vapour in proximity to the emulsion during the exposure, due to the variable airgap and increased atmospheric circulation above the edges of the plate in the open plateholder. Observations support this explanation; including the absence of the effect in plates exposed in a nitrogen flushed plateholder by Dawe and Metcalfe during 1980. The detailed investigation of background variations and correlations with telescopic and enviromental factors was taken up by A. W. Campbell as an M.Sc. project (Campbell 1981). The conclusions of Campbell's investigation confirmed the results described here.

Indications from this work and eye examination of plates indicate that the variations are due to changes in sensitivity and not variable fogging or other additive background variations. On this hypothesis the COSMOS measurement procedure - thresholding relative to local sky-intensity - should be insensitive to the background variations. This contradicted the results obtained with the COSMOS machine and further investigation of additive background effects was undertaken. Background

fog uniformity was investigated by measurement of fog levels in two situations; (a) around the edges and stepwedges of IIIaJ direct plates used in the investigation, (b) over the full extent of the three plates used in the calibration tests described below. Macbeth measures revealed no fog gradients in excess of 0.03D across the plates and in most cases no gradients were detected above the measurement errors ($\pm$ 0.015D), examples are given in Table 4.1. The large fog variations found by Metcalfe (1980 unpublished) were confined (as Metcalfe suggested) to plates exhibiting dark bands along one or more edges. This banding is caused by damage in transit, is readily identifiable and not present on any of the plates used in this investigation. Deep sky-limited plates with high background densities of at least 0.9D, (i.e. sky-background is at least 0.7D above plate fog) are little affected by the variations in fog described. These observations examined only the extreme edges of fully exposed plates, and it is possible some effect operates to produce fog variations in the centers of plates but this seems unlikely. However, considerable work has been undertaken on short exposure plates with COSMOS:- the sky-background is barely present on these plates, background densities of 0.3D with fog levels of >0.1D are typical - changes in the fog level of order 0.03D will significantly effect COSMOS measures, effectively causing Tc in equation 3.1 to vary with position. The large field effects observed in wide field photometric studies of such plates are quite consistent with this amplitude of fog variations. It is concluded that changes in background fog on UKSTU plates are generally small $\pm$0.015D and not a significant cause of field effects in COSMOS measures of deep sky-limited plates.

Field effects in COSMOS measures were also detected on the heavily vignetted AAT plates as well as the UKSTU hypered IIIaJ plates, both showed the same dependence on local sky-background. This suggested that

the COSMOS measuring procedure was not achieving its purpose for some as yet undetermined reason; attention was given to the interaction between the machine measurement procedure and the sky-background variations - Chapter 3. The results of this investigation were consistent with background variations due to sensitivity changes. The shape of the calibration curve, and variations in slope make little practical differerence when very low surface brightness objects are considered - i.e. faint stars and galaxies - as their component pixels are so close to the sky density. At brighter magnitudes changes in the shape of the calibration curve with plate position can result in large field effects; a direct check on the consistency of the calibration relation over an UKSTU plate was attempted.

4.3 Calibration Uniformity

To enable a direct empirical check on the sensitivity changes and possible changes in the shape of the calibration curve M. E. Sim of the UKSTU obtained three plates with twelve sets of sensitometer spots well distributed over the plates - all exposed similtaneously. Sim's experimental procedure was as follows; twelve attenuators of the type developed by Tritton (1981) for the new UKSTU twentyfive spot sensitometer were attached to a clear glass plate. The attenuators were carefully masked to prevent stray light reaching the plate during exposure. The array of attenuators together with the plate to be exposed were placed in a light box, exposed for fifteen seconds and then developed immediately. Although this investigation is unique it suffers from several severe limitations, the primary one being the short exposure time employed - 15s - which precludes the investigation of any photographic effects with a time dependence - e.g. reciprocity failure - that may be expected to vary during a long exposure. I am indebted to

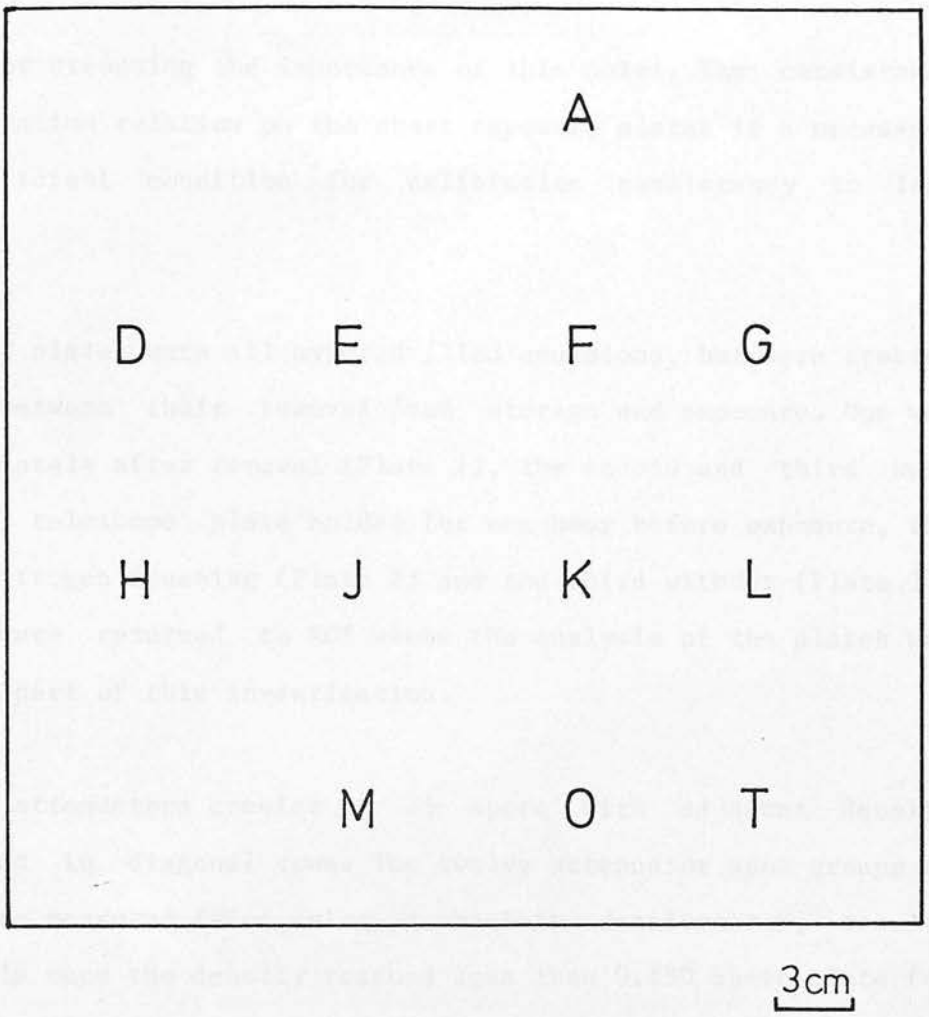


Fig. 4.3 The distribution of the attenuator spot groups on each of the three test plates described in the text. Letter designations are used to identify the attenuators - each of which is approximately 2cm on a side.

David Malin for stressing the importance of this point. The consistency of the calibration relation on the short exposure plates is a necessary but not sufficient condition for calibration consistency in long exposures.

The three test plates were all hypered IIIaJ emulsions, but were treated differently between their removal from storage and exposure. One was exposed immediately after removal (Plate 1), the second and third were kept in the telescope plate holder for one hour before exposure, the second with nitrogen flushing (Plate 2) and the third without (Plate 3). The plates were returned to ROE where the analysis of the plates was undertaken as part of this investigation.

The new UKSTU attenuators consist of 25 spots with adjacent density steps arranged in diagonal rows. The twelve attenuator spot groups on each plate were measured twice using a Macbeth densitometer, results were unreliable once the density reached less than 0.15D above plate fog and this data was discarded. Typically sixteen spots from each attenuator were measurable, in two cases the masking of the attenuators was not perfect and only twelve spots were measured. The attenuators themselves were measured three times with a Macbeth; density measures for the attenuators and each set of spots on the three plates were then available. Fig. 4.3 shows the arrangement of the attenuator spot groups on the plate, together with letter designations for each. Note the large separations between the spot groups which are well distributed over the plates.

RMS and systematic error estimates were made from the multiple measures: systematic shifts between attenuators and for measures over and between plates were $\leq$0.015D independent of density. The rms error for density

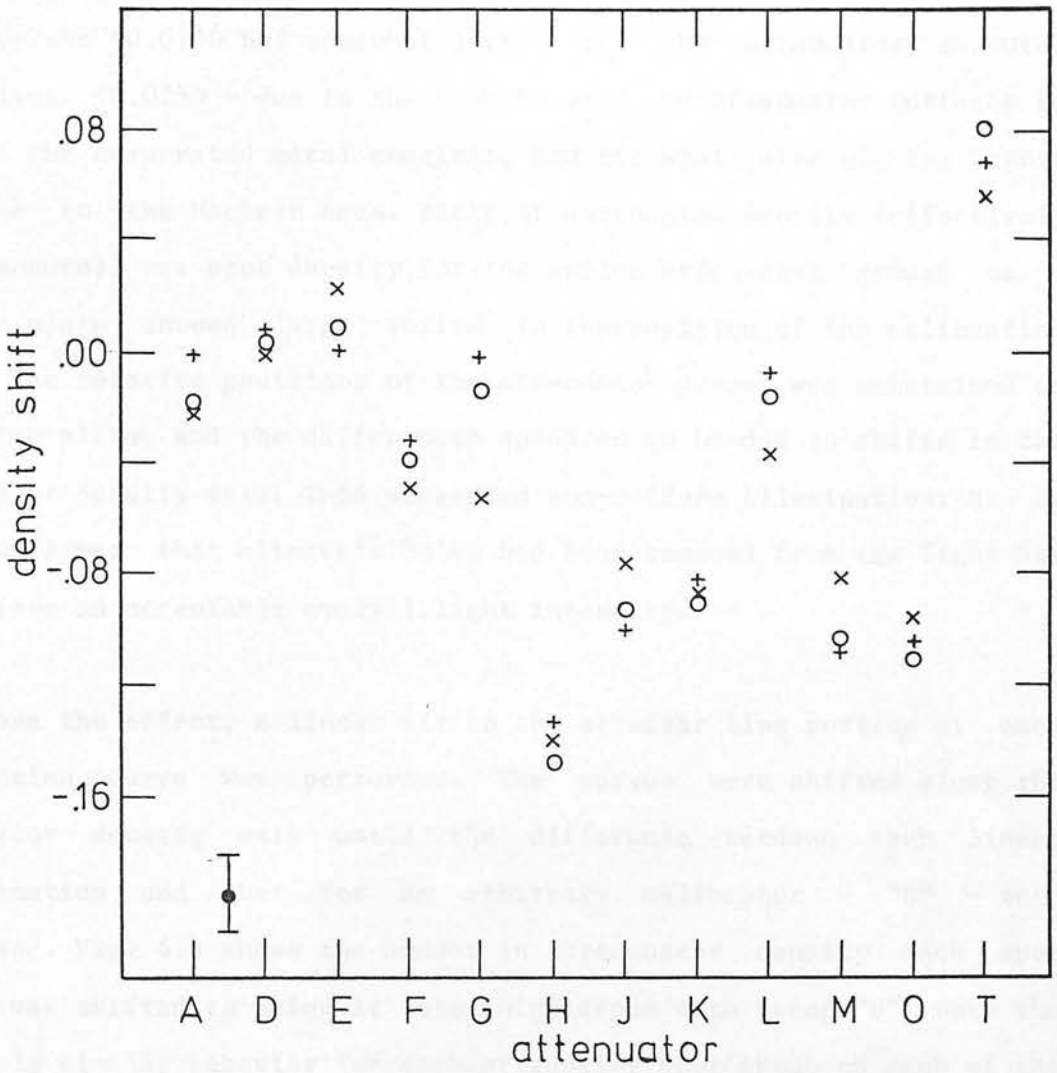


Fig. 4.4 The shift in the attenuator density axis necessary to bring each attenuator density curve into coincidence with attenuator "D". Plate 1 (+), Plate 2 (o), Plate 3 (×).

measures was $\leq$0.015D but somewhat larger for the attenuator measures themselves $\leq$0.025D - due to the need to mask the attenuator surfaces to protect the evaporated metal coatings, and the small size of the spots relative to the Macbeth beam. Plots of attenuator density (effectively Log(Exposure)) vs. spot density for the twelve attenuator groups on a single plate showed large shifts in the position of the calibration curve. The relative positions of the attenuator groups was maintained on all three plates and the differences appeared to be due to shifts in the attenuator density axis. This suggested non-uniform illumination: M. E. Sim confirmed that alternate bulbs had been removed from the light box to achieve an acceptable overall light intensity.

To remove the effect, a linear fit to the straight line portion of each calibration curve was performed. The curves were shifted along the attenuator density axis until the difference between each linear approximation and that for an arbitrary calibrator - "D" - were minimised. Fig. 4.4 shows the amount in attenuator density each spot group was shifted to bring it into coincidence with group "D". Note the extremely similar behavior for each attenuator spot group on each of the plates. The differences between plates are small and within the errors expected from the line fitting, and the assumption that the sensitivity change between plates for attenuator "D" is equal to the mean for the plates. Unless all the plates show identical sensitivity changes as a function of position these results rule out the possibility of global or local sensitivity changes of more than 2% in the regions of the plates occupied by the attenuators. Cubic spline curves with two knots were fitted to the resulting composite calibration curves. The fitting was performed using a program developed by S. Heathcote employing NAG spline fitting routines. The residuals from this mean relation were calculated for each attenuator spot group, and plotted against attenuator density

for each attenuator spot group. A strong correlation between the residuals for the same spot on the three different plates was noted, this correlation is readily explained by the nature of the measurement procedure; the slope of the attenuator spot group density vs. attenuator density - i.e. the gamma for the emulsion -, is 2 above a density of D=1. An error in the attenuator density results in a residual of twice the original error, as the same attenuator density measures are used for all three plates the errors are strongly correlated. Due to technical difficulties in measuring the attenuators - primarily the masking procedure necessary - it was not possible to achieve significantly improved accuracy. The correlation was removed by calculating the mean residual for each spot from all the plates and subtracting this from the individual plate residuals. Corrected results for Plate 2 as a function of attenuator spot density are shown in Fig. 4.5. A spot density of 0.1 corresponds to density 3 on the plate, and 1.2 to density 0.3 on the plate.

Remaining small systematic effects in the residuals are of the correct amplitude and qualitative form to be explained by the differential illumination operating over the area of a single attenuator spot group - only the mean effect was removed by the procedure described above. The arrangement of the spots means a gradient in residual along each diagonal row may be evident due to non-uniform illumination. The data was insufficient to determine a map of the incident light intensity and as the remaining systematic effects are small, no attempt has been made to apply a further correcton. The residual plots for all three plates show no evidence for changes in slope or shape of the calibration curves above measurement errors. This places an upper limit to the changes in calibration curve slope of 2% for the data under investigation. No detectable changes in slope of the linear portion of the calibration

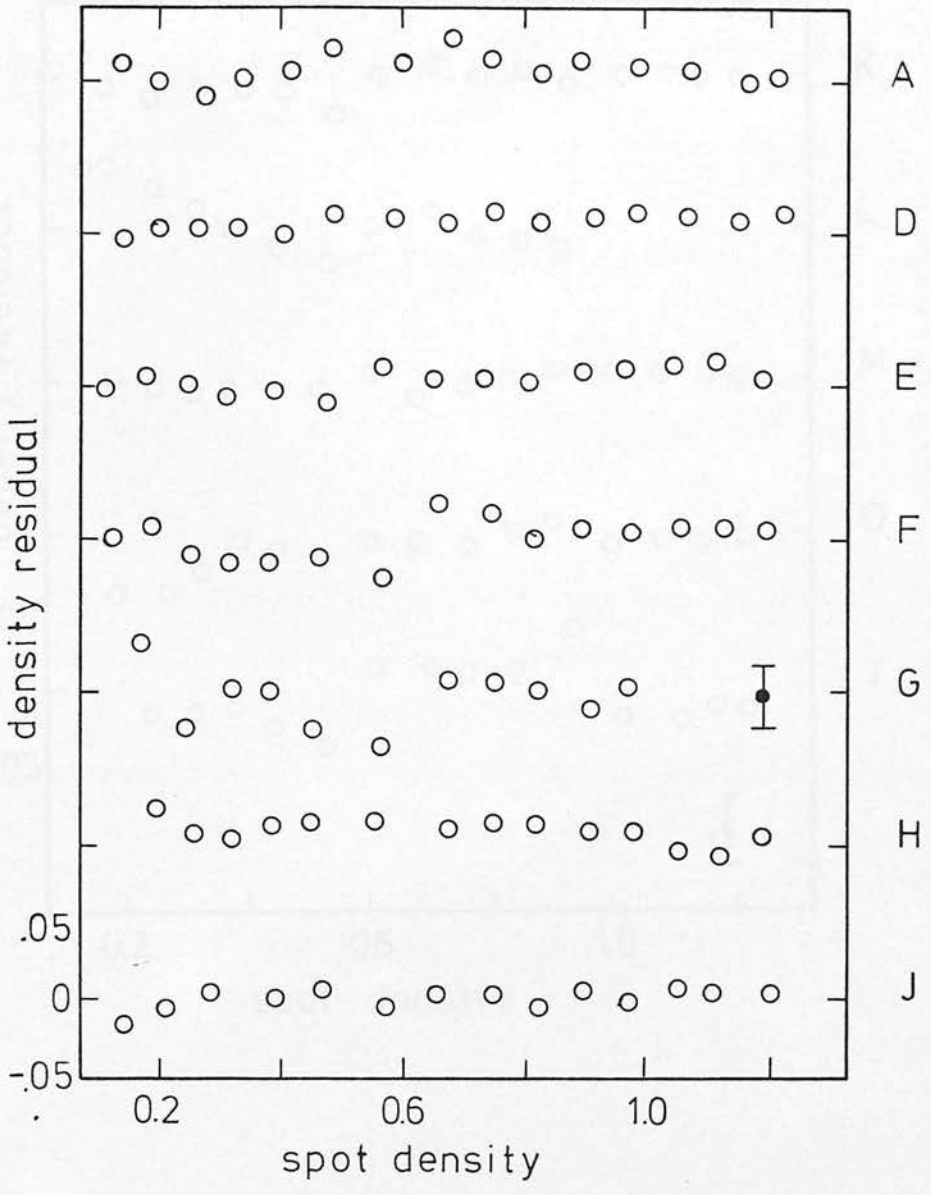


Fig. 4.5 The density residuals for the twelve attenuators on Plate 2 as a function of attenuator spot density. The data have been corrected for differential illumination effects and artificial correlations introduced in the spot density measures.

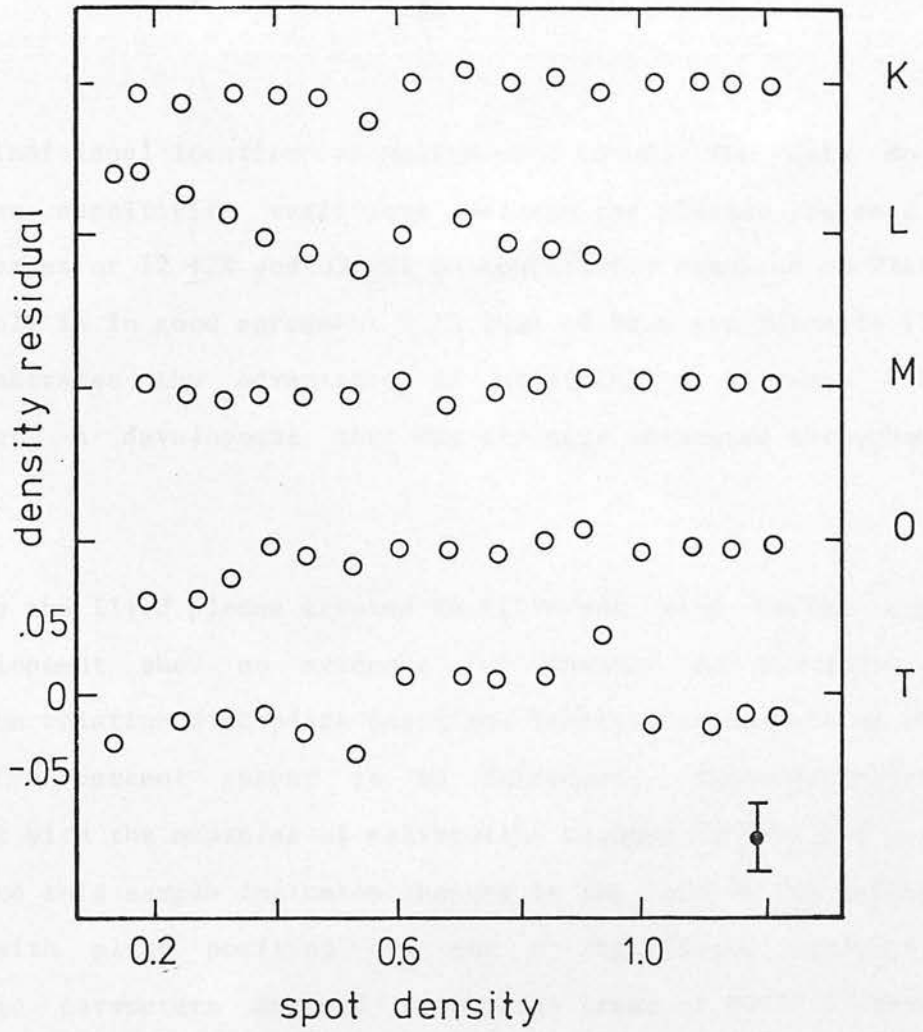


Fig. 4.5 (cont.)

curve or individual localised anomalies were found. The data do show large mean sensitivity variations between the plates; plates 2 and 3 show decreases of 12 $\pm$2% and 32 $\pm$2% in sensitivity compared to Plate 1. This result is in good agreement with that of Dawe and Metcalfe (1982), and demonstrates the advantages in utilising a nitrogen flushing plateholder, a development that was strongly advocated throughout the project.

In summary the IIIaJ plates treated in different ways before exposure and development show no evidence for changes in the shape of the calibration relation with plate position. Sensitivity variations of more than a few percent appear to be infrequent. The conclusions are consistent with the measures of calibration changes by Dawe and Metcalfe (1982), and this sample indicates changes in the form of the calibration relation with plate position are not a significant limitation on photometric parameters derived from large areas of UKSTU plates. This conclusion is qualified by the limited applicability of the data to long exposures.

4.4 Low Surface Brightness Emulsion Features

During eye classification of images on UKSTU plates an unusual feature of the emulsions was noticed; at very low surface brightness (lsb) large numbers of linear structures can be seen in the emulsions. The structures are characterised by very small cross-sections - 10-20 microns - but very large extents in one direction - 100-500 microns. The structures extend over many arcseconds and form sinuous elliptical shapes. The structure widths are smaller than the seeing disc and deeper plate material confirmed that the features are not threshold image detections. A high magnification reproduction (*10) of several of these

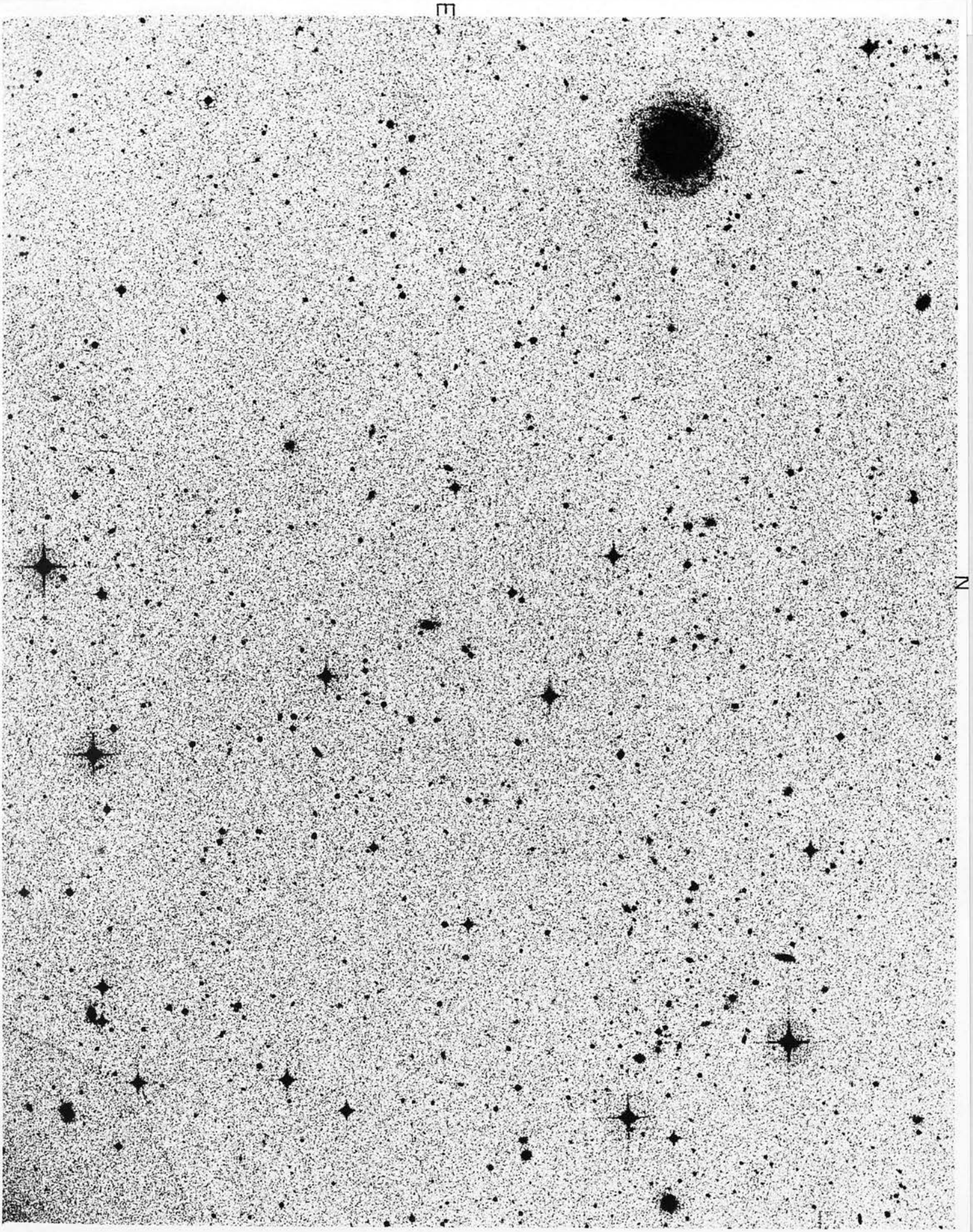
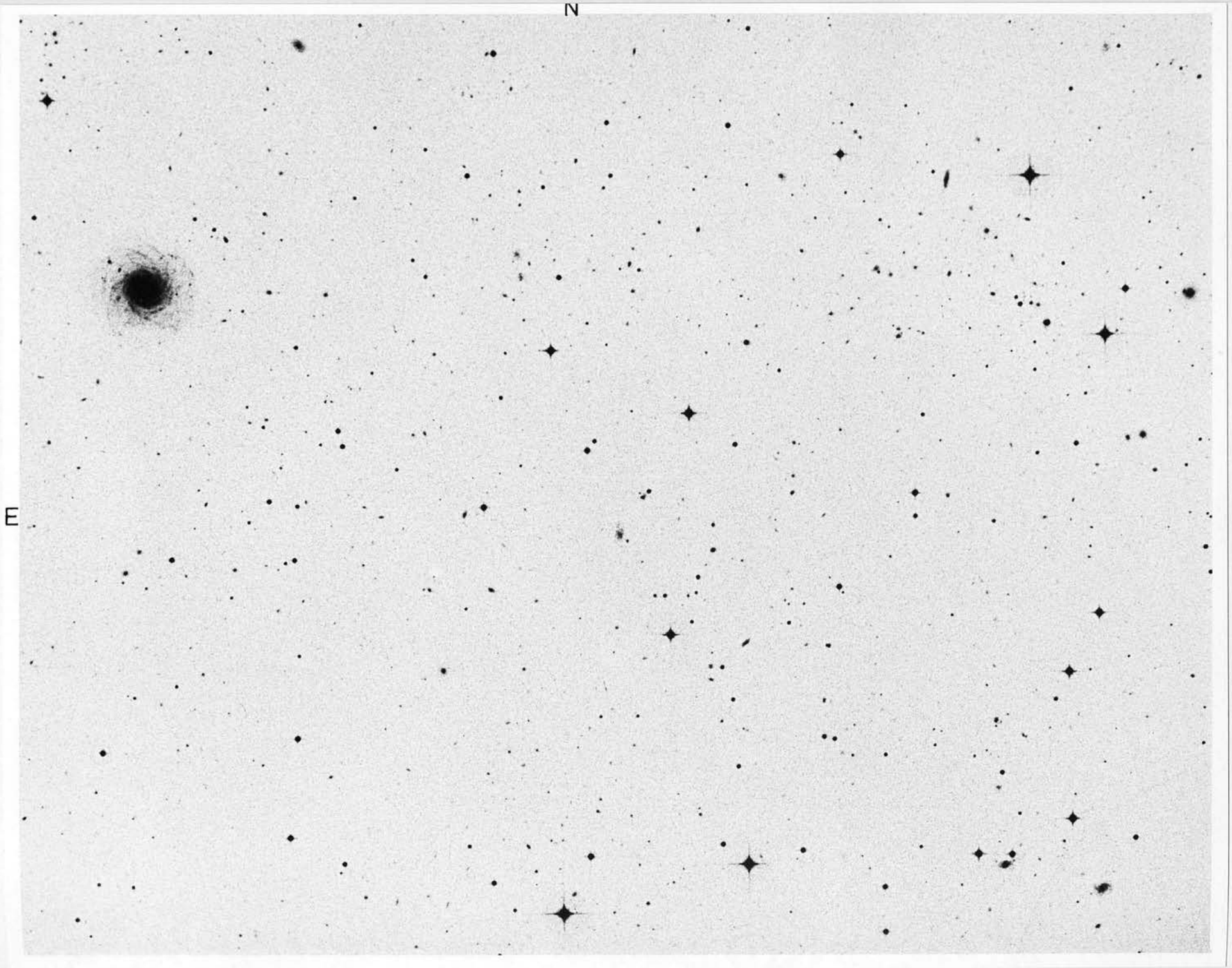


Fig. 4.6 A reproduction (10 times magnification) at normal and high contrast of a small area of plate J3721. Numerous sinuous sharp features can be discerned in the emulsion. Measuring from the south-east corner of the prints, particularly obvious features are evident at: 85:250, 242:187, 242:115 and a compact oval is present immediately to the north of the large galaxy - coordinates in millimeters.


N
E

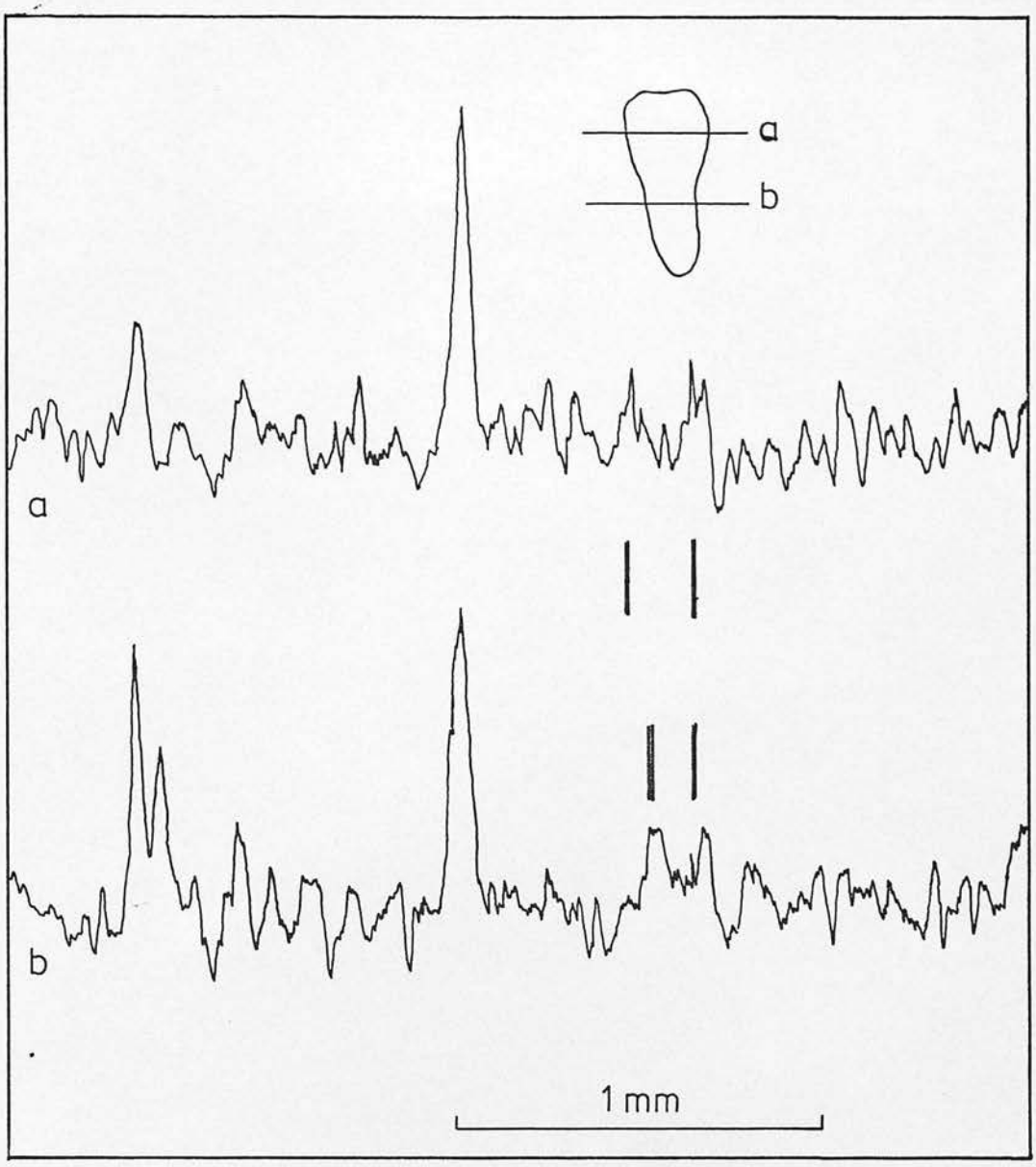


Fig. 4.7 Two Joyce Loebl tracings made across one of the 1sb features described in the text - slit dimensions 10*50 microns. The tracings are from plate J2652. The diagram at the top shows the positions of the two scans relative to the feature - scans were made approximately 80 microns apart. The vertical bars indicate the position of the edges of the elliptical feature in each of the scans. The large features in the scans are faint astronomical images.

structures is shown in Fig. 4.6a,b. The sharp linear nature of the features is further illustrated in Fig. 4.7 where two consecutive Joyce Loebl tracings perpendicular to the long axis of one of these features is shown - the slit dimensions were 10 * 50 microns. Note that the density of the features is not much greater from the general level of emulsion noise but that the structures show correlation over both scans. It is the correlation over hundreds of microns that is significant; the extent of the features is orders of magnitude greater than the dimensions of typical grain clumps - grain sizes on IIIaJ emulsions are expected to be of order 0.5 microns and grain clumping might be expected on scales of 10 microns. The features described here have nothing to do with grain clumping in the normally accepted sense of the term.

The intensity of the detectable structures above sky is of order 5-20%, the lower limit set by limitations of the eye in detecting lsb objects. The features are also visible in COSMOS mapping data but in general the structures are not detected as individual objects in COSMOS I.A.M. scans. However, whenever an astronomical image and one of the structures overlap, COSMOS combines the image pixels and the lsb structure pixels close to the image. Consequently the image aquires a spurious lsb component. This proves to be a major limitation in the classification of faint images and has apparently been ignored in previous work. Treatment of S/N for emulsions (Latham et al. 1976) and simulation of image classification techniques (Jarvis and Tyson 1981) essentially ignore this type of feature. The long chain like nature of the features makes them a serious problem for automated measurement; an image is contaminated whenever it crosses a chain. Sophisticated pattern recognition algorithms would be required to identify pixels affected by the features and remove their contribution from genuine image pixels.

Excellent agreement was found between objects eye classified as being contaminated by lsb features and those showing anomalous behavior in their COSMOS I.A.M. parameters. Inspection by eye, and examination of anomalous COSMOS images showed that image coincidence with one of the lsb features was a major noise source in the COSMOS data. The surface density of lsb features varies significantly from plate to plate but eye counts in two square centimeter areas on three J plates (J3390, J3721, J3406) indicate a typical surface density of 250-500 per square centimeter.

Discussions with members of the UKSTU revealed that few people were aware of the features, but they were easily seen when pointed out. The structures are evident on all types of astronomical emulsion, both AAT and POSS plates show large numbers of features - the frequency of occurrence is very high on the POSS prints. Copies of plates reproduce the structures present on the originals but a detailed comparison between copies and originals is difficult - the very different background densities of copies and originals makes lsb features much more obvious to the eye on copies. The presence of the features on several emulsions and different plate types rules out telescopic and hypering effects as possible causes. Their nature is strongly suggestive of some chemical process either in manufacture or development stages. Facilities to confirm this were not available and as the features already exist on the plates available for study, further investigation into the cause of the structures was not pursued.

4.5 Completeness Limits and Spurious Images

The optimum method for determining completeness limits and removing noise images is to pair together COSMOS measures of different plates of the region under investigation. This is not always possible due to the lack of suitable plate material, and available pairing programs are not yet sophisticated enough to deal with many plate combinations. Unless plates of the same colour and overall specification are used, severe practical problems arise when plate pairing is attempted - even for well matched plates the pairing becomes increasingly incomplete with fainter magnitude.

It is not possible to directly determine the completeness limits of the UKSTU plates without additional, superior quality data, however, an attempt to determine the completeness limits and noise frequencies for COSMOS measures of UKSTU plates through comparison with AAT plates was made. COSMOS output of small areas - 3 square centimeters - of UKSTU plates was used to determine the distribution of faint images - the faint limit being set by the COSMOS area cut of ten 8 micron pixels. In practice away from bright (m_B<16) images COSMOS is phenomenally good at detecting faint images without generating noise images; examination of 40,000 images in the image classification program revealed only one image recorded by COSMOS for which a corresponding feature on the plate could not be seen - the contribution to noise images in the general field due to COSMOS may be ignored. Systematic variations in detection criteria may also be ignored due to the small areas of plate examined; only COSMOS random errors and thresholding detection method need be taken into account when considering the completeness results. The completeness limit of the COSMOS measurement procedure had already been determined from examination of multiple COSMOS measures of single plates

- Chapter 3, and results for the plates of this investigation are given in Table 7.5. The COSMOS measurement process does not introduce significant numbers of noise images brighter than cosmos-magnitude zero, corresponding to an image surface density of >5000 per square degree on IIIaJ plates measured at 10% threshold cuts.

Noise Images: At very faint levels - $m_J \sim 22$ - on the IIIaJ plates noise images detected by COSMOS surrounding objects of intermediate brightness - $19 < m_J < 15$ - are the predominant source of spurious images. This is a consequence of the COSMOS background following technique and the occurrence of faint diffraction spikes and lsb outer regions of galaxies on the plates. This noise contamination becomes evident on IIIaJ plates measured at a 10% threshold cut in the size range 15-25 eight micron pixels. The detection of noise images purely due to fluctuation in the emulsions becomes evident around the 15 pixel level, but is still remarkably small. Table 4.2 gives example counts of noise images on IIIaJ and IIaD plates, in all cases deep AAT plates were used to determine the reality of faint images. The frequency of noise images is higher on copy plates as a small percentage of noise images are introduced in the copying procedure, but in no case examined was the noise level on copies raised by more than factor of two compared to the originals. The concentration of noise images around brighter images has severe effects on small scale clustering statistics and a realistic limit for use of the data is 30 pixels. For typical IIIaJs the noise frequency in the COSMOS data is $\leq 3\%$ to a number density of 5000 objects per square degree, corresponding to a magnitude of $m_J \sim 22$. A limit for IIaD plates is $\leq 3\%$ to 2000 objects per square degree - $m_V \sim 20.2$. The detection of large numbers of spurious images to these number densities is clearly not a significant problem. The noise frequencies quoted can be entirely ascribed to limitations in the plate material and are not

due to limitations imposed by COSMOS.

Table 4.2

Plate	Total images	Noise images	Image density	Magnitude limit
V3475	712	24	2300	20.3
V6619	710	10	2300	20.3
J3406	568	10	5000	22.0
J4445	568	6	5000	22.0

Plates	Total images	Completeness (i)	Completeness (ii)	Completeness (iii)
V3475/ V6619	712	90 (87%)	66 (91%)	95±2%
J3406/ J4445	568	39 (93%)	29 (95%)	98±2%

(i) all images not in both samples, and corresponding percentage completeness

(ii) astronomical images not in both samples

(iii) estimate of completeness allowing for random errors due to COSMOS

(a) image densities are per square degree

Completeness Limits: Data from multiple measures of the same plate give an estimate of the rms scatter in image parameters due to COSMOS - e.g. Fig. 3.15. Multiple measures of different plates give an estimate of the combined plate/machine error in the image parameters. Completeness of the UKSTU plates was considered in terms of number density by examining the same images on different plates - allowance being made for the random errors introduced by COSMOS. As a check, rms error estimates were combined with an extrapolation of the number counts on plates between $18<m_J<21$ to model the expected distribution for fainter magnitudes. Allowance was made for the loss of images from the sample due to the

COSMOS thresholding procedure - the figures quoted here do not include very lsb objects not detected by COSMOS because of the high threshold cut employed - 25.0-25.5 magnitudes per square arcsecond. The results of both methods agree well and indicate the plates become incomplete (>10%) at number densities per square degree of 6000 - m_J >22 - for COSMOS measures at 10% threshold cuts, of A grade, sky-limited, IIIaJ plates at high galactic latitude. This is in agreement with an eye comparison of the data of Carter (1980), from the AAT plate J1555 and UKSTU IIIaJ plates of the same region (J4445, J3406, J757). An identical procedure was performed for the V plates used in the investigation, giving a completeness limit (>90%) of 2500 objects per square degree ($m_V \sim 20.5$) on A grade high galactic latitude plates measured at a 10% threshold cut - example counts are given in Table 4.2. The data relating to COSMOS in Chapter 3 and UKSTU plates in this Section demonstrate that incompleteness due to large random or systematic errors between COSMOS measures of high quality UKSTU J and V plates is not a significant problem to number densities of 5000 and 2000 objects per square degree respectively. Although severe systematic effects occur due to the COSMOS measurement process the use of image number density to define the sample limits ensures a high degree (>90%) of consistency between samples from different UKSTU plates.

4.6 Summary

The existence of large sky-background variations (5-10% in relative intensity) on many hypered IIIaJ plates has been established. The properties of the variations are consistent with emulsion sensitivity changes as a function of plate position. Variations in the slope of the calibration curve on short exposure UKSTU plates have been shown to be small ($\leq$2%) and not a significant factor in limiting the precision of

astronomical data. Fog variations are found to be small ($\pm$0.015D), and in the case of deep sky-limited (D$\geq$1) plates not a limiting factor in automated measurement.

A source of noise at low surface brightness levels on photographic plates has been noted - apparently for the first time, the features result in a large increase in the noise affecting machine measured parameters of faint images. The origin of the features probably lies in the manfacturing or development processes. Examination of UKSTU, and deeper AAT plates shows that COSMOS object samples from high quality J survey plates are complete (>90%) to >5000 objects per square degree or $m_J \sim 22$. Noise images due to fluctuations in the emulsions are not a significant source of contamination - $\leq$3% - brighter than this limit.

5 Image Classification

5.1 Introduction

Quantitative analysis of an object catalogue for galaxy clustering requires some form of image classication; to improve the S/N of the clustering and to provide a companion stellar sample as a control. Initially classification of COSMOS data was attempted using techniques described by MacGillivray (1976) and Corwin (1981), data already classified using these techniques was also examined. The results were disappointing, with overall success rates much lower than claimed, combined with large position dependent variations in classifier success. The positional dependent variations were due to the parameter shifts described in Chapter 3. The large changes in classification success with plate position were confirmed by Corwin, who also found the overall success rates in small areas of plate were lower than expected from earlier work (Corwin 1981 and private communication). Attempts to classify objects over large areas using COSMOS measures showing significant parameter shifts was not attempted because of their complex nature. The remainder of this Chapter considers data where significant parameter shifts were not present.

Inspection of plate material by eye showed that the reliable assignment of images to well defined catagories - stars and galaxies for instance - was very difficult at faint magnitudes. Extreme difficulty was encountered fainter than m_J=21 on UKSTU IIIaJ plates limited at m_J=23, and considerable difficulty was experienced as bright as m_J =20.5, similar limitations with respect to the limiting plate magnitude were found on B, V, R and I plates - i.e. classification was extremely difficult within two magnitudes of the plate limit. This observation was

inconsistent when compared to some claims in the literature, where success rates for classification stratagies are quoted to much fainter levels with respect to the plate limits (e.g. Jarvis and Tyson 1981 and Peterson et al. 1979). Classification success rates are generally evaluated using eye examination of the same plate material as available to the automated classifier. Frequently no external verification is applied to automated techniques, consistency between eye and machine classification does not ensure either is able to classify objects. This apparently simple consideration does not seem to be generally appreciated.

The claimed ability to reliably eye classify objects near the plate limit of some workers contrasts with the study of Brown (1978). Brown found significant errors in his eye classification when deeper plate material was used to establish classifications for objects at the limit of shallower plates. To resolve the discrepancies evident in the literature and determine the reason for the low success rates of automated classification initially observed, a program of eye classification using very deep plate material was undertaken. This established a data catalogue against which the COSMOS classifications from shallower plate material could be tested. This appears to be the first occasion such an external check of the reliability of automated machine image classification on this scale has been undertaken.

5.2 Data

COSMOS measures of UKSTU B, V, R, I and J plates were used as the basis for the investigation. The plate material used is listed in Table 5.1. The control data for the COSMOS measures consisted of film and glass copies of accepted A grade J survey plates, together with other A grade

UKSTU J and R plates - in addition AAT J/R plate pairs were used. The AAT plates had limiting magnitudes 1.5-2 magnitudes fainter than the corrsponding UKSTU plates, coupled with a much larger plate scale, which is advantageous for image classification.

Table 5.1

Plate number	Plate grade	Images classified	Magnitude limit	Field designation
V2499 *	A	1500	20.5	MTF
V5421 *	A	2000	20.5	MTF
V3475 *	AH	2600	20.5	SGP
V5511 *	A	1100	20.5	SGP
V6608 *	A	1100	20.5	SGP
V6619 *	A	1100	20.5	SGP
J3721 *	AE	900	21.5	SGP
R4676 *	AE	900	20.0	SGP
I6523 *	BE	900	18.5	SGP
I6427 *	A	900	18.5	SGP
J757 *	BSP1	7000	21.0	F349
J3406 *	BP1	1500	21.5	F349
J4445 *	BPIT4	1000	21.5	F349
UJ7149 *	BID	2500	21.5	F349
J3001	A	1600	20.5	HERCULES CLUSTER
R5793	A	1600	19.7	HERCULES CLUSTER
		AAT Plates		
J1555				SGP
J1554				MTF
R1565				MTF
J2055				MTF
R2054				MTF
R1790				SGP
R1996				SGP
J1888				SGP

(a) magnitude limits are given in their respective plate bands. Errors $\pm$0.2 magnitudes

(b) * denotes plates for which AAT plates were employed for verification.

Multiple plates in more than one colour were always used when establishing the eye classifications. Ellipse plots from COSMOS I.A.M. measures of areas of plate to a given cosmos-magnitude limit were used as finding charts for the images to be classified. A total of 30,000

objects were classified; each object was examined on at least three different plates, including at least one with a plate limit a minimum of two magnitudes fainter than the object being classified. This number of object classifications was necessary because of the range of plate types and quality, and the need to determine accurate statistics on classification success rates and frequency of object types. Images were classified into one of seven catagories: (1) single stars, (2) multiple stars, (3) multiple objects (stellar component dominant), (4) multiple objects (non-stellar component dominant), (5) single non-stellar objects, (6) multiple non-stellar objects and (7) noise. The resulting catalogue of classifications is vastly superior in quality to any described in the literature, and enables a far more rigorous test of image classification stratagies than possible previously. Previous checks of COSMOS and other machine data have often been performed using only a few hundred images, many close to the plate limit where the eye classifications are unreliable.

5.3 Number Counts

Of great practical importance for determining the success of image classification is the ratio of stars to galaxies (S/G) in a sample. Over the magnitude range $14 \leq m \leq 24$ the ratio S/G varies by four orders of magnitude. A fixed percentage error in the classification of one type of object can have dramatically different effects on the contamination in the second type of object sample. At faint magnitudes this has resulted in disagreement over the stellar number counts (Jarvis and Tyson 1981, Kron 1980, Koo and Kron 1982), due to the small value of the ratio S/G. A smaller effect in the opposite sense occurs for the data in this thesis; the differential number counts of stars and galaxies are approximately equal at $m_J=20.5$. Brighter than this, where many of the

samples discussed in Chapter 9 are drawn, stars outnumber galaxies. A fixed percentage error in the classification of stars thus translates into a larger percentage contamination of the galaxy sample. Values of two to three times the stellar misclassification rate are found for contamination in the brighter - $m_J \leq 20.5$ - galaxy subsamples. For galaxy samples with contamination constrained to $\leq 10\%$, stellar classification accuracy of $\geq 95\%$ are necessary. The importance of this effect was evident in checks of classified COSMOS data at the start of this investigation: up to 25% of the total galaxy samples - $m_J \leq 20.5$ - consisted of bright - $m_J \leq 18.5$ - misclassified stars.

5.4 COSMOS Parameters

The COSMOS I.A.M. parameters available are listed in Table 3.2. The use of I.A.M. output clearly limits the number of possible discrimination stratagies - any technique involving the use of profile information for example is not feasible. To keep the technique as general as possible it is assumed throughout that no prior information is fed into the classifier procedure (e.g. position of galaxy clusters, ratio of stars to galaxies), this reduces the number of parameters listed in Table 3.2 that might be useful as discriminators. A range of potential discriminators and the parameters from which they are derived is given in Table 5.2. A basic ordering parameter is required to separate images according to apparent size or magnitude. In some previous work with COSMOS, Log(area) had been used but the COSMOS I.A.M. output had improved significantly and cosmos-magnitude was found to be much less susceptible to noise, and therefore a better ordering parameter. The quantities listed in Table 5.2 were tested as image classifiers using subsamples of 1000-2000 images from different plates.

Table 5.2

Parameter designation	Parameter description	
A	cosmos-magnitude	-2.5*Log(Σ(I-Isky)/Isky)
B	peak surface brightness	-2.5*Log((Imax-Isky)/Isky)
C	ellipticity (unit weighted)	1-Au/Bu
D	ellipticity (intensity weighted)	1-Ai/Bi
E	second moment (unit weighted)	π*Au*Bu/area
F	second moment (intensity weighted)	π*Ai*Bi/area
G	Log of image area	Log(pixels)
H	Gaussian dispersion	σ=sqrt((area/2)*Ln(Imax/Ithresh))
I	magnitude residual from fitted gaussian	ΣI - Igaussian
J	occupation index	(ΔXmax *ΔYmax)/area
K	relative centroid shift	sqrt((Xu-Xi) + (Yu-Yi))/R*

(a) parameter H is the dispersion of a gaussian intensity profile with the same peak and threshold intensity as the COSMOS image.

(b) parameter I is the magnitude difference between the integrated intensity of a gaussian image with central intensity Imax and a dispersion chosen to approximate the seeing on the plate. The gaussian intensity is calculated by integrating to radius R*.

(c) R* is the equivelant radius defined as R* = sqrt(area/π).

5.5 Results

5.5.1 Eye Classification; the following conclusions from the eye examination of images are relevant; (a) Properties of images of the same type - e.g. single stars - vary dramatically over the magnitude range present on UKSTU plates. A bright fourteenth magnitude star with a saturated core, prominant diffraction spikes, surrounding halation and an area of several thousand 8 micron pixels bears little relation to a twentyfirst magnitude star consisting of fifty 8 micron pixels with no obvious structure, and particularly susceptible to contamination from

several noise sources. Simple invariant parameters to describe images with such different properties do not exist in the I.A.M. output, and a basic ordering parameter, such as cosmos-magnitude must be used - i.e. at least a two dimensional classifier is required.

(b) The COSMOS I.A.M. output merges a substantial fraction of images - 5-10% - on UKSTU plates at the percentage threshold cuts used - 7-10% of sky-intensity. The fraction is strongly dependent on image density and hence galactic latitude, but for latitudes $b \geq 40^{\circ}$, apparent magnitude $m_J \leq 21$, 5-10% of the images are merged, this sets an upper limit to the classification success rates - Section 5.5.2.

(c) The preliminary difficulties in classifying images within two magnitudes of the plate limit were confirmed: classification of objects by eye on shallow plates followed by classification on deeper plates showed that classification was no better than 85% successful at $m_J = 21$. Fainter than this a rapid decline in success rate occurs, with no useful classification being possible for $m_J > 22$. Comparison of 362 galaxies (limited at $m_J \sim 20.5$) classified on plate J3721 with reliable identifications from three AAT plates - Table 5.1 - revealed only two galaxies whose classification was uncertain, and both of these were assigned to the galaxy class. If this field region is typical then a significant population of "compact" objects in galaxy samples to $m_J \sim 20.5$ does not exist. The classification errors are in good agreement with the data of Brown (1978). The authors eye classification was verified by comparison with classifications made by H. G. Corwin and S. B. Tritton and excellent agreement was found between classifications: eye classifications used in this investigation do not suffer from significant differences with those of other workers.

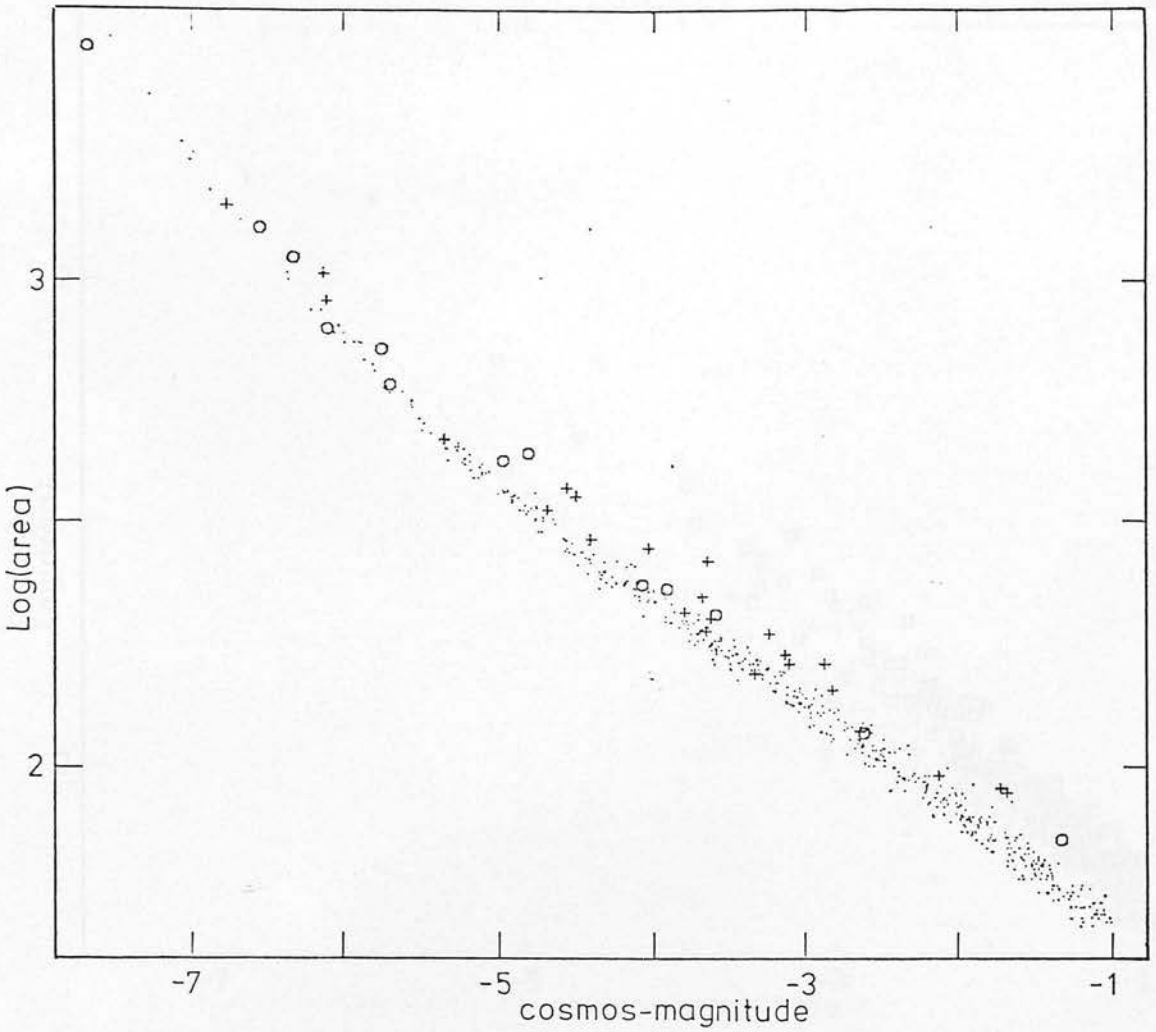


Fig. 5.1a,b The distribution of 1100 images from plate V3475 in cosmos-magnitude vs. Log(area) space. The parameters for the identical group of images are plotted in all further diagrams of data from V3475. The symbol types are maintained throughout Chapter 5 and Appendix 1: single stars (·), multiple stars (O), star-galaxy (+), galaxy-star (×), single galaxy (□), multiple galaxy (◇). For each pair of figures, part "a" contains the distribution of stellar dominated images, and part "b" the distribution of diffuse images with the single stellar objects plotted for reference. The stellar magnitude limit of the sample is approximately $m_J \sim 20.5$.

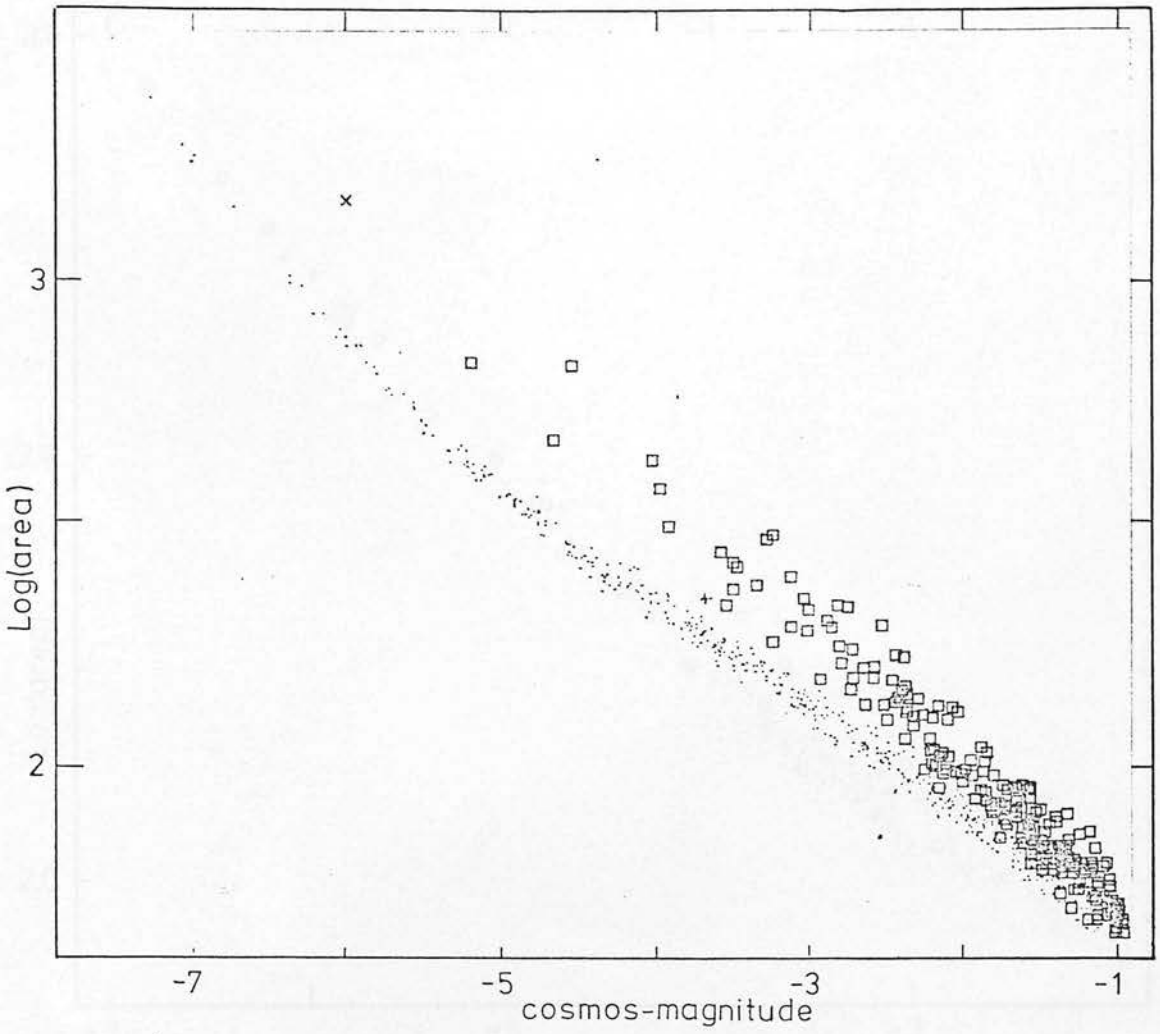


Fig. 5.1b

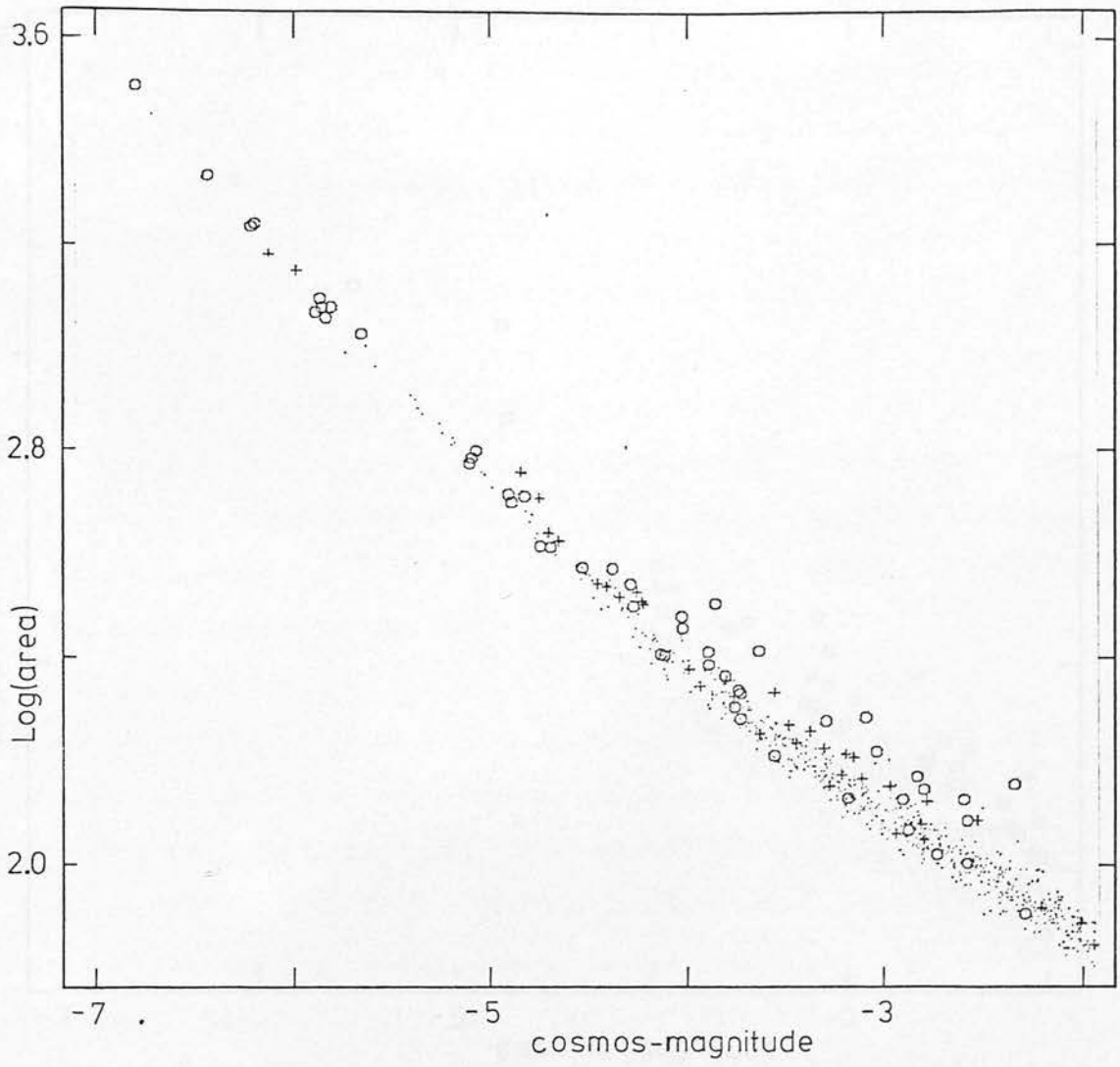


Fig. 5.2a,b The distribution of 750 images from plate J3001 in cosmos-magnitude vs. Log(area) space. The stellar magnitude limit for the sample is approximately m_J=20.5.

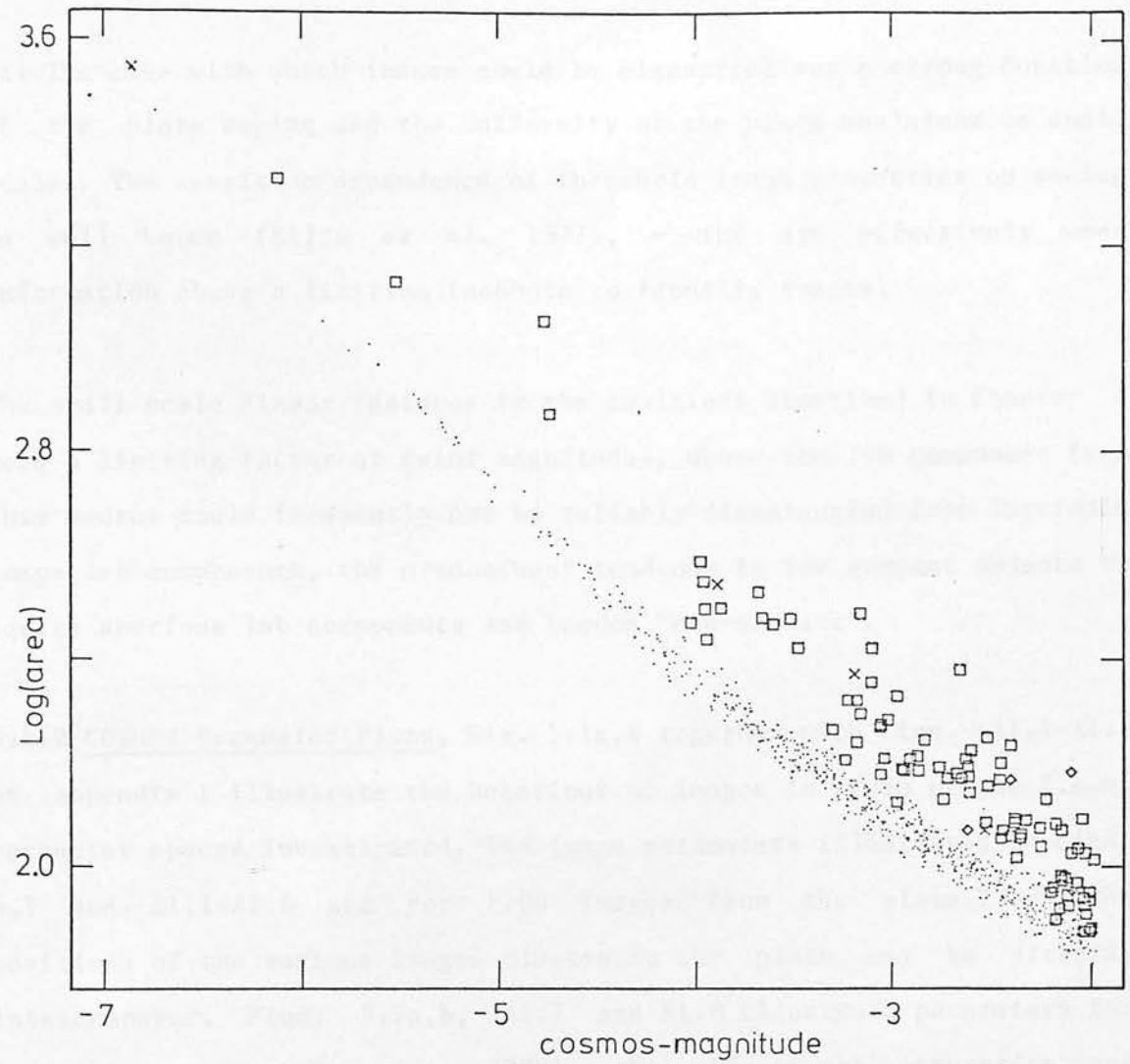


Fig. 5.2b

(d) The ease with which images could be classified was a strong function of the plate seeing and the uniformity of the plate emulsions on small scales. The sensitive dependence of threshold image properties on seeing is well known (Ellis et al. 1977), - the eye effectively uses information above a limiting isophote to identify images.

The small scale linear features in the emulsions described in Chapter 4 were a limiting factor at faint magnitudes, where the lsb component from this source could frequently not be reliably disentangled from intrinsic image lsb components, the predominant tendency is for compact objects to aquire spurious lsb components and become "non-stellar".

5.5.2 COSMOS Parameter Plots; Fig. 5.1a,b together with Figs. A1.1-A1.6 of Appendix 1 illustrate the behaviour of images in seven of the I.A.M. parameter spaces investigated. The image parameters illustrated in Figs. 5.1 and A1.1-A1.6 are for 1100 images from the plate V3475, the positions of the various images classes in the plots may be directly intercompared. Figs. 5.2a,b, A1.7 and A1.8 illustrate parameters for bright images from the plate J3001, the effects of saturation and transmission quantisation are far more obvious in these plots, which are typical of those from deep sky-limited IIIaJ plates. The faint stellar magnitude limit of the V3475 and J3001 data is m_J=20.5. The following general conclusions were drawn from the detailed examination of the behaviour of eye classified images in COSMOS I.A.M. parameter spaces; (a) Within one to one and a half cosmos-magnitudes of the sky brightness virtually no discrimination between different image types is possible in the parameter plots. In this region the position of an image is primarily determined by plate and measurement noise - the entire range of parameter space being spanned by "identical" stellar objects. The machine noise is large at these magnitudes - Fig. 3.15 - and

combined with the intrinsic plate noise prevents discrimination at these levels. A more sophisticated software analysis procedure may improve the situation.

(b) None of the I.A.M. parameters provide good separation at both bright and faint magnitudes, confirming the conclusions drawn from the eye classification. The dynamic range in cosmos-magnitude of each discriminator is strongly dependent on plate type via the severity of saturation effects (either plate or machine induced) for the brighter images - cf. V3475 and J3001 data.

(c) Simple shape parameters such as unit or intensity weighted ellipticity are of little use for image separation because of the wide spread in intrinsic galaxy shapes. Further as images become fainter, the point spread function and the plate noise result in a strongly peaked distribution at intermediate ellipticity and no useful separation is possible.

(d) It is not possible to separate single and multiple images using COSMOS I.A.M. parameters without removing a substantial fraction of the galaxy sample (>10%) which show asymmetric intensity profiles. The systematic rejection of this type of galaxy is not desirable, and progress in this field will require more detailed information on image profiles. The frequency of occurrence of multiple stellar images sets an upper limit to the success of image classification with the parameters listed in Table 3.2. Figs. A1.2 and A1.4 illustrate two parameters that provide the best discrimination between single and multiple images, but note the overlap of galaxies and multiple images.

(e) The positions of objects relative to the stellar sequence in many parameter plots are strongly correlated. I.A.M. parameters are derived from one image intensity profile and are not independent. This is particularly important when the effects of non-poisson noise sources are considered. The lsb features described in Chapter 4 and other emulsion flaws seriously modify the intensity profiles of images, and several parameters are detrimentally affected, usually in the same sense: a contaminated star becomes a definite "galaxy" in several parameter spaces - combining the classifications from different spaces does not result in an improved classification. The limitations imposed by non-poisson noise sources are not always appreciated, and it is not possible to "beat down the noise" by simply combining different parameters. In most cases working with one parameter that is least susceptible to noise is better than degrading the classifications by using several parameters, one of which is more susceptible to noise. For example, the PSB criteria works well at faint magnitudes because of the high S/N of the central image pixels, and low susceptibility to contamination by lsb features, which in contrast, make Log(area) of little use in the same magnitude range on poor quality plates. This effect is illustrated in Figs. 5.3a,b and 5.4a,b, where the PSB and Log(area) parameters for the same set of images from plate R4676 are shown; the rather noisy background on this plate makes Log(area) of little use between -2< cosmos-magnitude <-1 although PSB still provides a reasonable separation.

(f) For the reasons outlined in this section the method for achieving the most successful image classification stratagy employs different discrimination parameters over different magnitude ranges, rather than simply combining parameters over wide magnitude ranges - this is a new approach as far as COSMOS data is concerned.

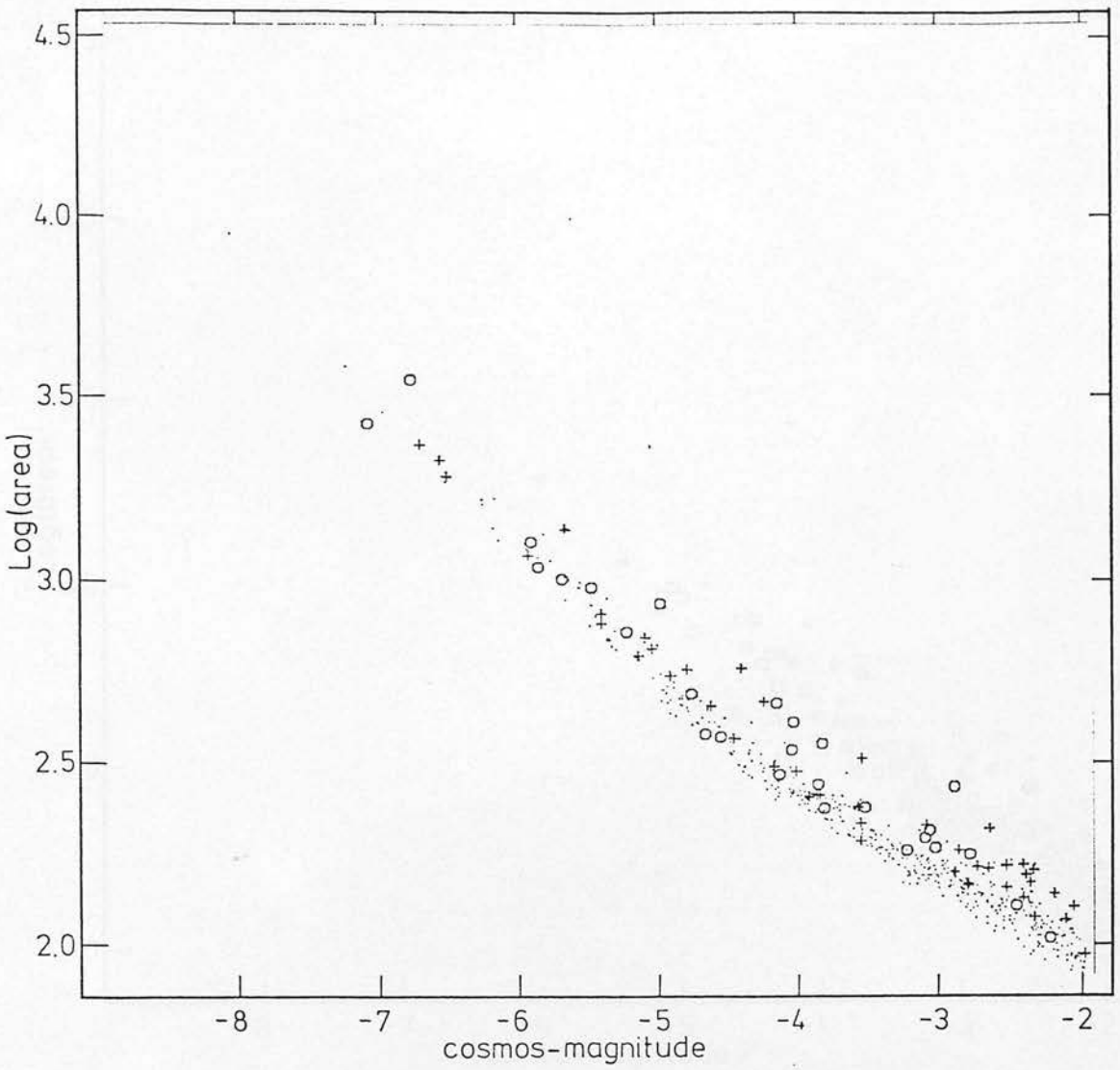


Fig. 5.3a,b The distribution of 900 images from plate R4676 in cosmos-magnitude vs. Log(area) space.

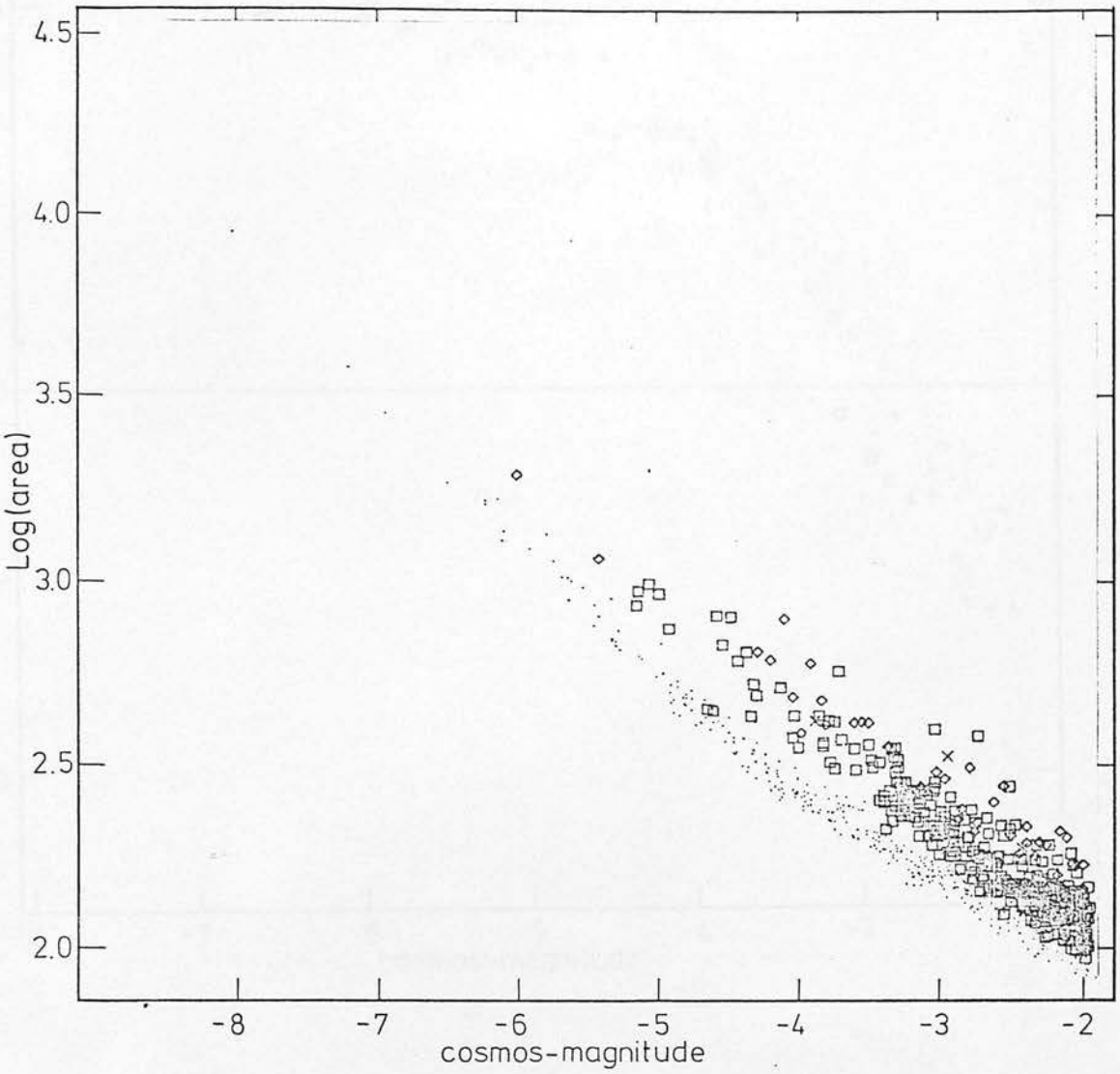


Fig. 5.3b

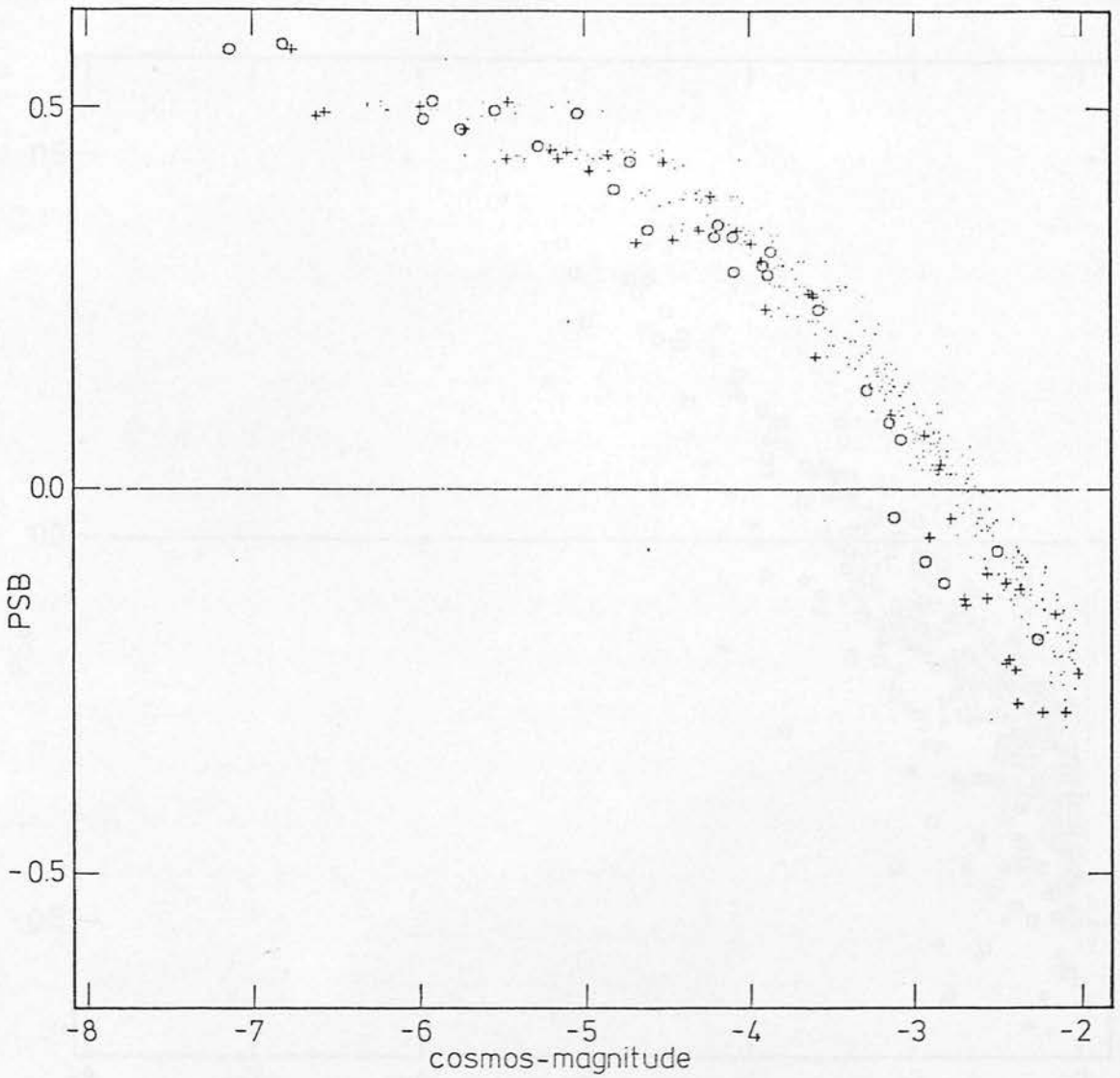


Fig. 5.4a,b The distribution of 900 images from plate R4676 in cosmos-magnitude vs. peak surface brightness space. Compare the relative distribution of stellar and non-stellar objects at faint magnitudes with that in Fig. 5.3.

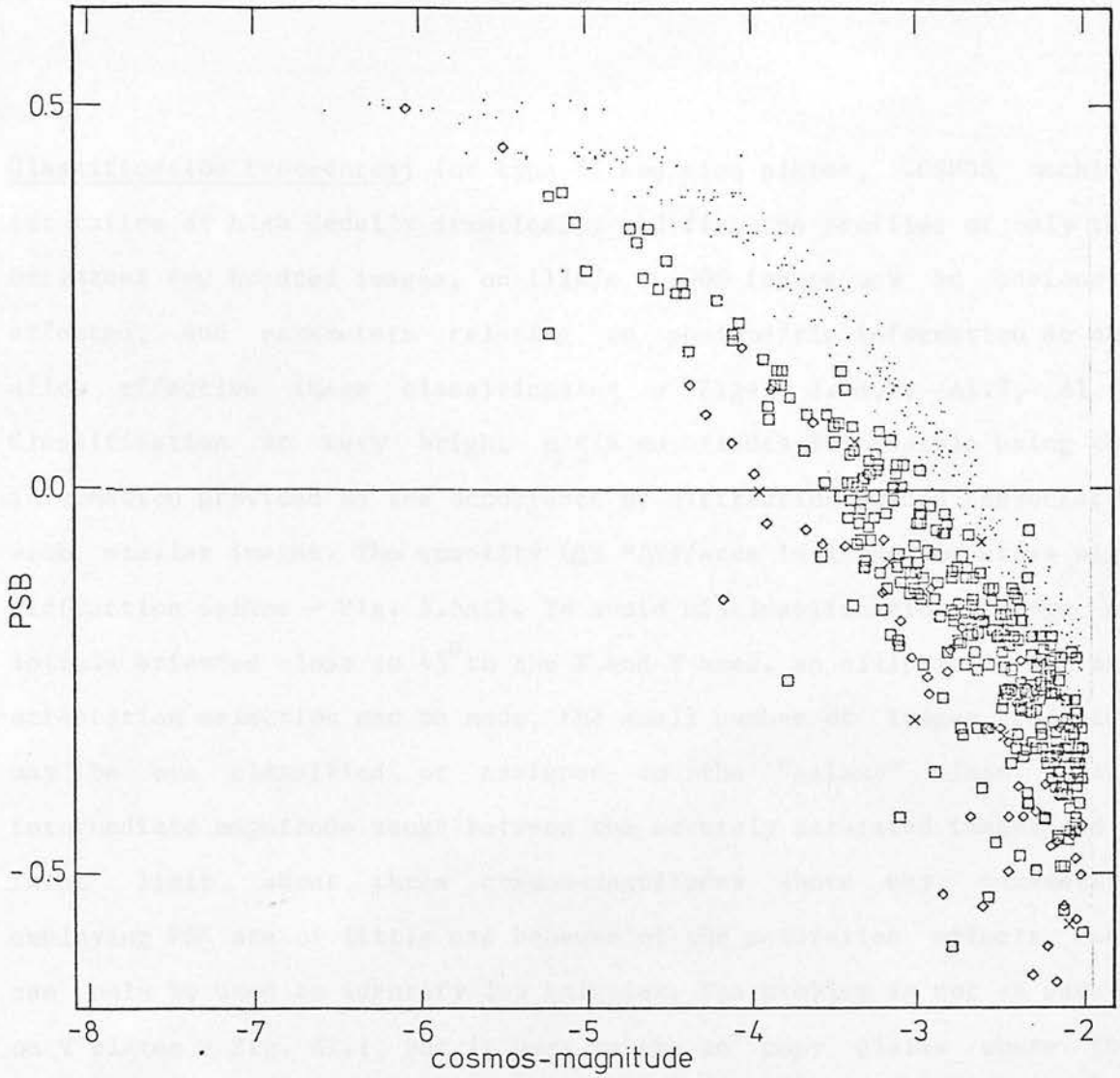


Fig. 5.4b

Classification Procedures; for type II emulsion plates, COSMOS machine saturation at high density drastically modifies the profiles of only the brightest few hundred images, on IIIaJs 10,000 images may be seriously effected, and parameters relating to photometric information do not allow effective image classification - Figs. 5.4a,b, A1.7, A1.8. Classification at very bright $m_J \leq 16$ magnitudes is possible using the information provided by the occurrence of diffraction spikes associated with stellar images. The quantity ($\Delta X * \Delta Y$)/area is large for stars with diffraction spikes - Fig. 5.5a,b. To avoid misclassification of edge on spirals oriented close to 45° to the X and Y axes, an ellipticity cut and orientation selection can be made, the small number of images rejected may be eye classified or assigned to the "galaxy" class. In the intermediate magnitude range between the severely saturated images and a faint limit about three cosmos-magnitudes above sky, parameters employing PSB are of little use because of the saturation effects, and can only be used to identify lsb galaxies. The problem is not as severe on V plates - Fig. A1.1, but is very severe on copy plates where the plate saturation combines with that from COSMOS. Parameters utilising information from the lsb portions of images work most effectively in this region as they are independent of saturation; Log(area) is the most effective discriminator, Ai * Bi/area (IWSM) is effective a magnitude or two below the onset of visible diffraction spikes, and there is a strong correlation between the positions of objects in the two parameter spaces. At magnitudes fainter than two cosmos-magnitudes above sky the S/N of the lsb parameters decreases rapidly; Signal is proportional to image area (i.e. $\propto r^2$) and Noise to circumference (i.e. $\propto r$) - fainter objects are more reliably classified using PSB, due to the higher S/N of the central intensity pixels. Below one cosmos-magnitude above sky little separation is possible using any single, or combination of

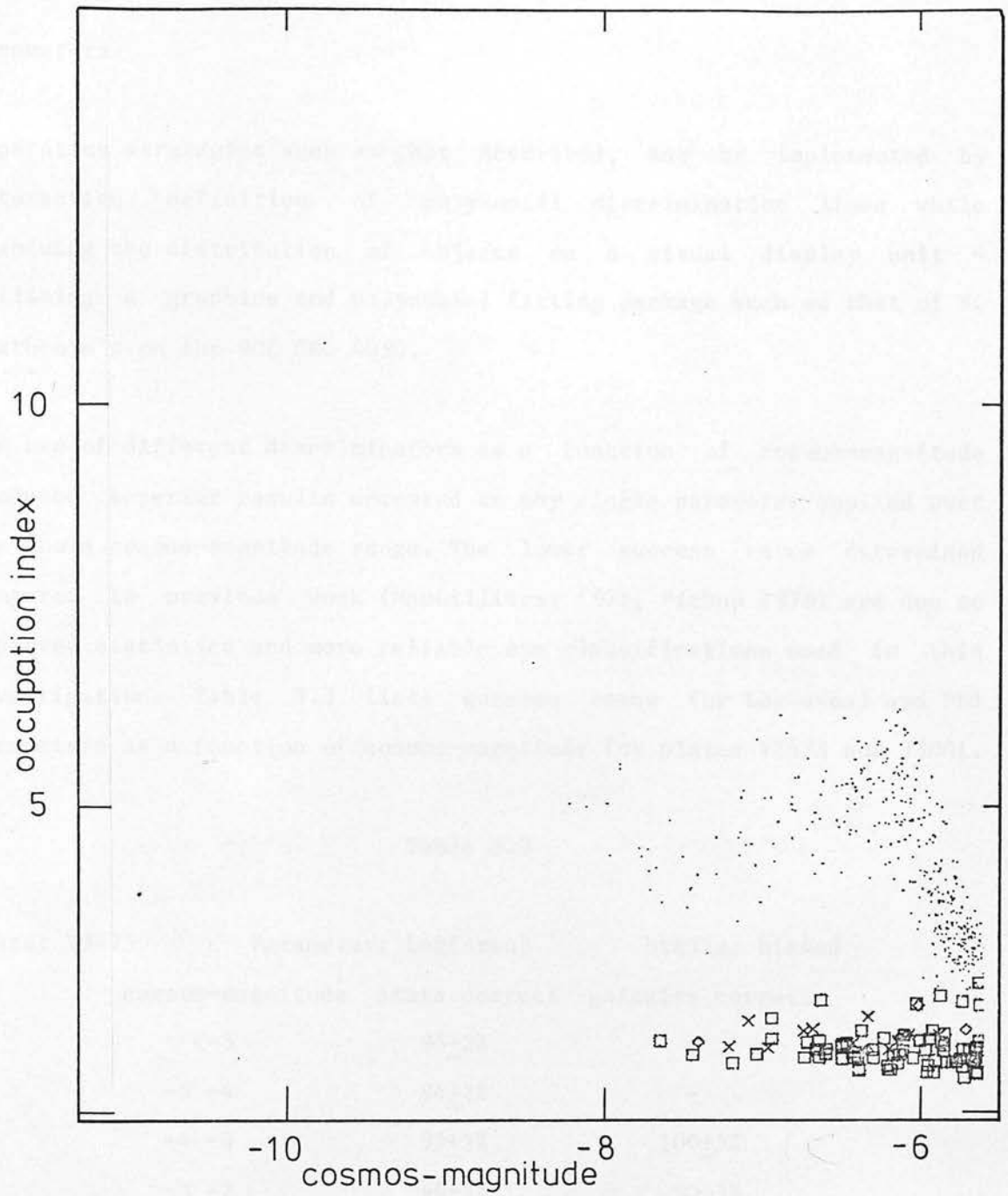


Fig. 5.5 The distribution of 500 images from plate J3001 in cosmos-magnitude vs. occupation-index space. An ellipticity cut to exclude images with intensity weighted ellipticities greater than 0.3 and orientations within 15 degrees of the plate diagonals has been applied.

parameters.

Separation stratagies such as that described, may be implemented by interactive definition of polynomial discrimination lines while examining the distribution of objects on a visual display unit - utilising a graphics and polynomial fitting package such as that of S. Heathcote's on the ROE GEC 4090.

The use of different discriminators as a function of cosmos-magnitude produced superior results compared to any single parameter applied over the whole cosmos-magnitude range. The lower success rates determined compared to previous work (MacGillivray 1976, Pickup 1979) are due to improved statistics and more reliable eye classifications used in this investigation. Table 5.3 lists success rates for Log(area) and PSB parameters as a function of cosmos-magnitude for plates V3475 and J3001.

Table 5.3

Plate: V3475 Parameter: Log(area) Stellar biased

cosmos-magnitude	stars correct	galaxies correct
<-5	95±2%	-
-5 -4	94±2%	-
-4 -3	95±2%	100±5%
-3 -2	96±2%	90±3%
-2 -1	97±1%	70±2%

Practical effect:

Galaxy sample limited at	-2	8% lost	25% contamination
	-1	22% lost	13% contamination

Table 5.3 (cont)

Plate: V3475 Parameter: Log(area) Galaxy biassed

cosmos-magnitude	stars correct	galaxies correct
<-5	95±2%	-
-5 -4	94±2%	-
-4 -3	95±2%	100±5%
-3 -2	95±2%	97±3%
-2 -1	31±2%	98±2%

Practical effect:

Galaxy sample limited at	-2	2% lost	27% contamination
	-1	2% lost	61% contamination

Intermediate discrimination

Practical effect:

Galaxy sample limited at	-2	5% lost	26% contamination
	-1	9% lost	23% contamination

Plate: V3475 Parameter: PSB Stellar biased

cosmos-magnitude	stars correct	galaxies correct
<-5	100±2%	-
-5 -4	99±2%	-
-4 -3	94±2%	94±4%
-3 -2	92±1%	96±3%
-2 -1	97±1%	75±2%

Plate: V3475 Parameter: PSB Galaxy biased

cosmos-magnitude	stars correct	galaxies correct
<-5	100±2%	-
-5 -4	99±2%	-
-4 -3	94±2%	94±4%
-3 -2	92±1%	97±3%
-2 -1	40±2%	96±2%

Table 5.3 (cont)

Plate: J3001 Parameter: Log(area) Stellar biased

cosmos-magnitude	stars correct	galaxies correct
<-5	100±3%	-
-5 -4	97±3%	-
-4 -3	93±3%	97±5%
-3 -2	88±2%	90±3%

Practical effect:

Galaxy sample limited at -2 8% lost 37% contamination

Plate: J3001 Parameter: Log(area) Galaxy biased

cosmos-magnitude	stars correct	galaxies correct
<-5	100±3%	-
-5 -4	97±3%	-
-4 -3	93±3%	97±5%
-3 -2	82±2%	97±3%

Practical effect:

Galaxy sample limited at -2 3% lost 47% contamination

Plate: J3001 Parameter: PSB Stellar biased

cosmos-magnitude	stars correct	galaxies correct
<-5	100±3%	-
-5 -4	100±3%	-
-4 -3	95±3%	65±4%
-3 -2	97±2%	62±4%

Practical effect:

Galaxy sample limited at -2 37% lost 8% contamination

Table 5.3 (cont)

Plate: J3001 Parameter: PSB Galaxy biased

cosmos-magnitude	stars correct	galaxies correct
<-5	100±3%	-
-5 -4	99±3%	-
-4 -3	85±3%	79±3%
-3 -2	68±2%	87±3%

Practical effect:

Galaxy sample limited at -2 16% lost 100% contamination

(a) formal poisson errors assumed throughout.

5.6 Automated Classification

Substantial improvements over interactive definition of discrimination lines between image types are not possible using automated techniques; the limitations on success rates being set by the occurrence of multiple images and noise at faint magnitudes. However comparable success rates can be obtained using an automated classifier with the advantage of being able to combine classifications from different parameters and plates. An automated classifier was developed to enable rapid processing of large numbers of COSMOS measures.

Images from COSMOS I.A.M. scans were binned into arrays representing the image number density in parameter spaces - images near bright stars and other anomalous regions were excluded. The position of the modal stellar locus was determined for each parameter using non-linear least squares fitting: a cross-section of parameter space for each cosmos-magnitude interval was taken, and a gaussian fitted to the image number density

profile. The fit was only performed to the images in the compact - i.e. stellar - region of the parameter space to prevent contamination by non-stellar images. A polynomial fit was made to the positions of the stellar locus - i.e. the centers of the fitted gaussians - to eliminate noise in the individual determinations. This polynomial defines the position of the modal stellar sequence. The stellar dispersion was calculated by fitting a second polynomial to the dispersion of the fitted gaussians. Technically the determination of the dispersion should be performed perpendicular to the local tangent to the modal polynomial - experiment showed that no significant improvement in discrimination was obtained using this procedure, and for ease of computation, sigma was determined at fixed cosmos-magnitude. The use of the data on the compact side of the stellar locus ensures that the locus position and stellar sequence dispersion are not significantly biased by the presence of non-stellar images until considerable overlap of stellar and non-stellar images occurs. In this region, as indicated earlier no separation is possible in any case.

Using the equation for the stellar locus and the values of the stellar dispersion at each cosmos-magnitude, every object in a parameter space can be assigned a distance from the stellar locus in terms of the local dispersion. Parameter classifications can be easily combined in this "normalised" classification space - different parameter classifications or classifications from different plates may be weighted as desired. The adopted discrimination limit may be chosen at fixed dispersion; constant fraction of stars correctly classified with an increasing contamination from galaxies at fainter magnitudes, or the fixed dispersion criteria may be modified by a second function that reduces the dispersion criteria with magnitude ensuring that all galaxies are assigned to the "galaxy" class with increasing stellar contamination at fainter

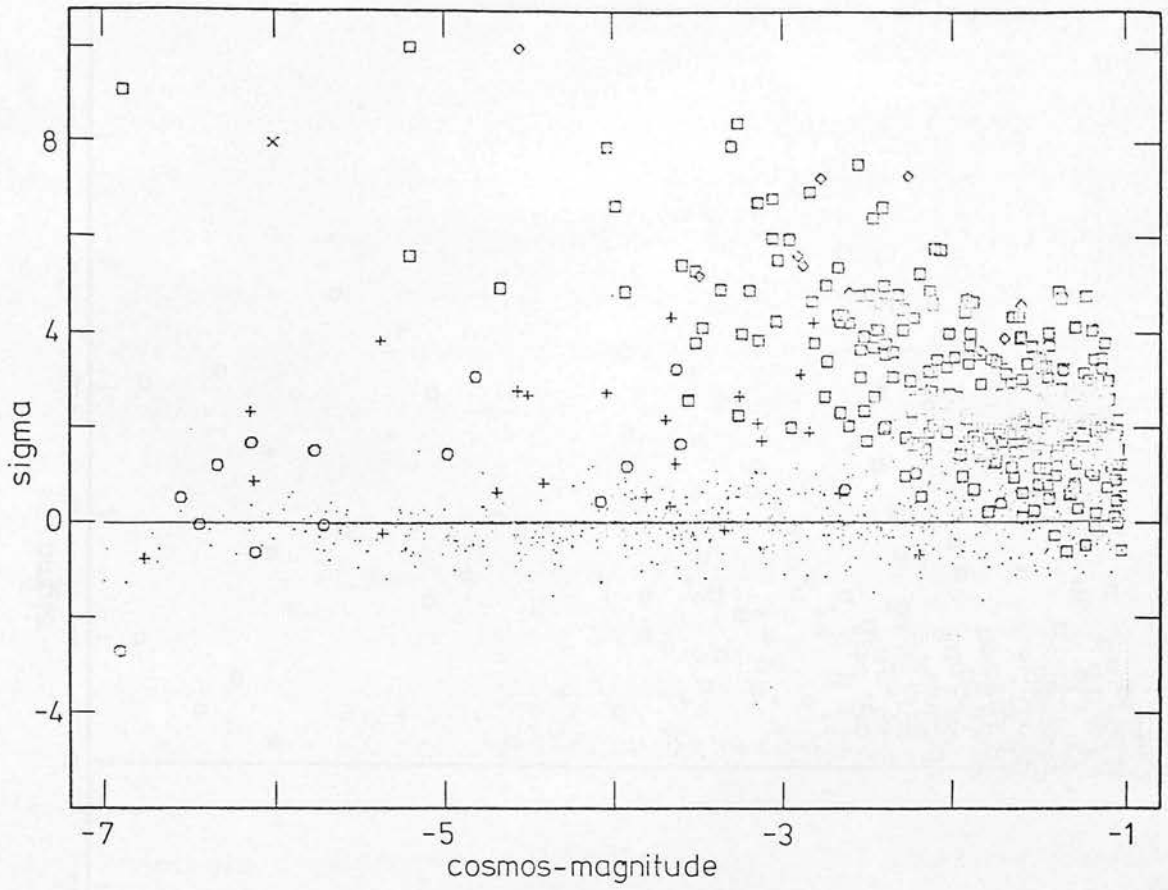


Fig. 5.6 The distribution of 1100 images from the plate V3475 in cosmos-magnitude vs. distance from stellar locus, where the distance is expressed in number of gaussian standard deviations calculated from cosmos-magnitude vs. Log(area) space. The dispersion was calculated using the automated classifier described in the text - compare with Fig. 5.1.

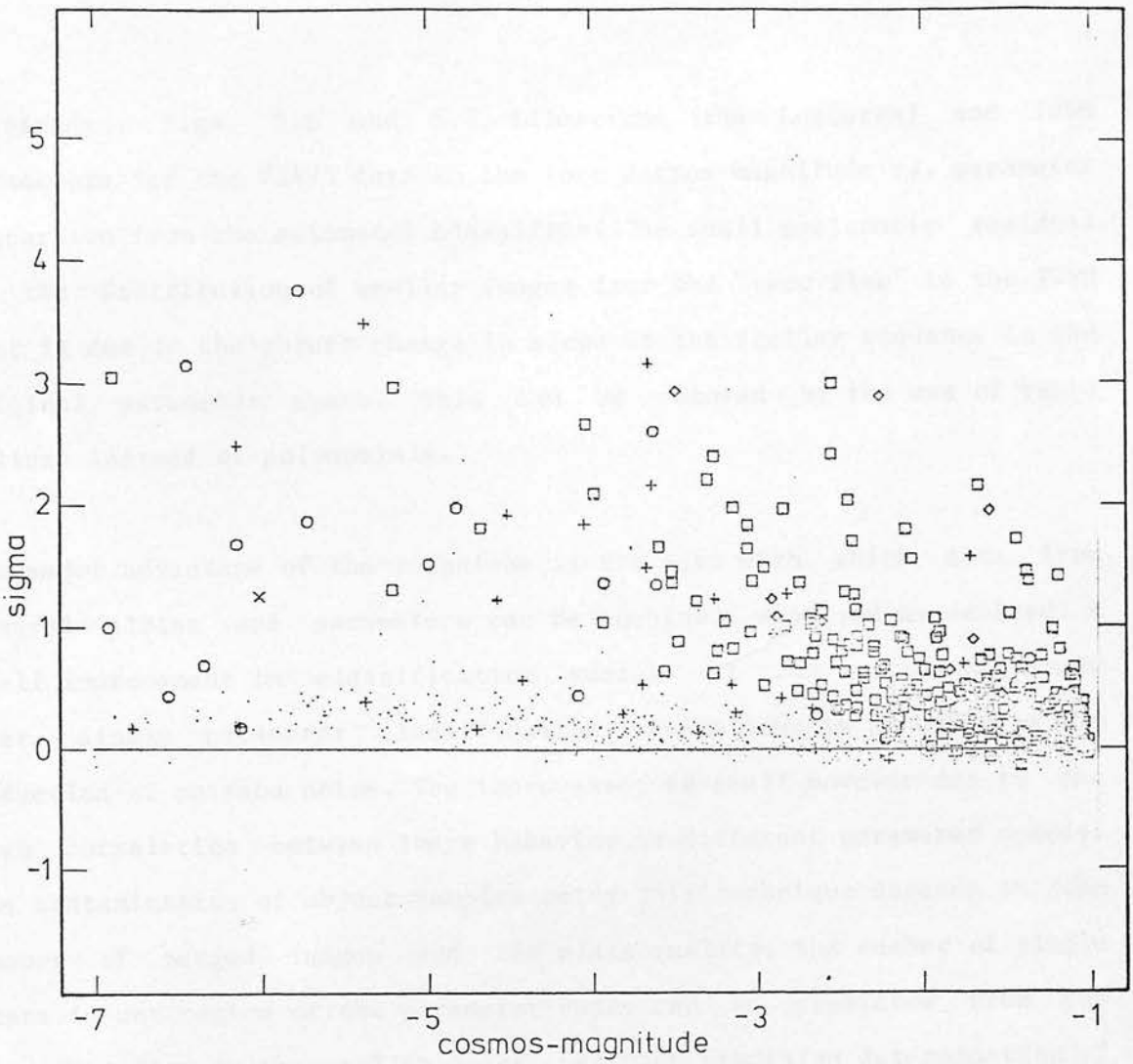


Fig. 5.7 The distribution of 1100 images from the plate V3475 in cosmos-magnitude vs. distance from stellar locus, where the distance is expressed in number of gaussian standard deviations calculated from cosmos-magnitude vs. intensity weighted second moment space. The dispersion was calculated using the automated classifier described in the text - compare with Fig. A1.3.

magnitudes. Figs. 5.6 and 5.7 illustrate the Log(area) and IWSM parameters for the V3475 data in the form cosmos-magnitude vs. parameter dispersion from the automated classifier. The small systematic residual of the distribution of stellar images from the "zero line" in the IWSM plot is due to the abrupt change in slope of the stellar sequence in the original parameter space. This can be removed by the use of cubic splines instead of polynomials.

The major advantage of the technique is the ease with which data from several plates and parameters can be combined, weighted as desired. A small improvement in classification success at the cosmos-magnitude where single parameter classification breaks down is possible by the reduction of poisson noise. The improvement is small however due to the high correlation between image behavior in different parameter spaces. The contamination of object samples using this technique depends on the number of merged images and the plate quality; the number of single stars in any region of the parameter space can be predicted from the gaussian fits to the stellar locus, but high precision determination of success rates - i.e. to better than five percent - can not be achieved without eye classification of images on the plate. The automated technique breaks down where stellar and non-stellar images show substantial overlap in parameter space, and in its present form would not be suitable for classification at faint magnitudes - $m_J > 22.5$ - where the number of stars is small and a stellar sequence could not be reliably determined.

5.7 Discusion

Classification of images using COSMOS data has been attempted by MacGillivray (1976), Pickup (1979), McNally (1979) and Corwin (1981):

MacGillivray and Pickup used a PSB critria, but note that the normalisation by sky-intensity used by Pickup and here is necessary if the parameter is to remain invariant to plate sensitivity changes. Pickup found the technique broke down at m_J=21.5, and MacGillivray found a similar limit. Corwin's data obtained over the same period as the early part of this investigation is in exact agreement with the data presented here: the conclusion of this investigation suggests a lower limit for image classification (integral probability of classification success $\geq$90%) of $m_J \sim 20.5$ on high quality IIIaJ plates, this limit is brighter than determined by MacGillivray (1976) and Pickup (1979), but is explicable in terms of the superior eye classifications of this investigation. The differential classification success declines rapidly fainter than $m_J \sim 20.5$ on J plates and $m_V \sim 18.5$ on V plates. The variation in classification success due to image number density and different quality plates can not be stressed too highly, however, given the large number of plates examined in this investigation it is unlikely that significantly better results than those for plate V3475 in Table 5.3 will be obtained with this type of COSMOS data. COSMOS transmission saturation was found to be a serious problem which does not appear to have occurred at the time of MacGillivray or Pickup's investigations, this could be due to their poor statistics, or substantial changes in the COSMOS configuration may have resulted in saturation occurring at fainter magnitudes. The inability to identify a technique capable of discriminating between single and multiple images confirmed the work of McNally (1979) and Stobie (private communication). The frequency of merged images found was larger than determined by MacGillivray and Pickup but the same as McNally, this is probably due to the improved statistics of this investigation.

Comparison of this work with techniques implemented on other measuring machines is difficult; the noise properties of the machine, type of image profile information provided and plate qualities all vary considerably. The most elegant technique, utilising information over a wide range of surface brightness is that developed by the Oxford galaxy group (e.g. Carter and Godwin 1979, Godwin et al. 1982). The technique involves the construction of a point spread function (psf) from eye identified stellar profiles: the difference in slope between the composite psf and the profile of each image provides the discrimination parameter. The success rates claimed for this technique, and plots of images in this parameter space (Carter and Godwin 1979; Fig. 2, Godwin et al. 1982; Figs. 6 and 7) are in excellent agreement with the conclusions reached here. The Oxford group's experience with eye classification is also similar, with extreme difficulty being encountered within two magnitudes of the plate limit (Godwin 1981, private communication). Peterson et al. (1979) effectively used the same technique without the normalisation of the slope parameter using the psf. Their claim to reliably eye and machine classify images fainter than m_J=23.5 disagrees with data presented here, and the parameter diagram (Peterson et al. 1979; Fig. 1) does not suggest their data justifies this claim.

Kron (1980) used the image pixel intensities weighted by the inverse square of the radial pixel distance, this weights the central pixels strongly. The technique is similar to the PSB parameter used here, and the quoted success rates are comparable - Kron's technique can be extended to much brighter limits than with COSMOS due to the different saturation properties of the PDS. Kron does however use his automated classifier well below the magnitude limit where it breaks down.

The most sophisticated techniques applied to image classification are due to Sebok (1979) and Jarvis and Tyson (1981; henceforth JT): both have made attempts to use as much information as possible, the resulting techniques are inherently complex but the claimed success rates are high. JT use higher order moments of the image intensity distribution in attempting separation. JT's experience with information from the outer lsb portions of images contradicts the results from the COSMOS Log(area) parameter for example, where noise was found to be a limiting factor before some other parameters were significantly degraded. Kron (1980) also remarks that the image moments are strongly correlated and little was gained by going to higher orders. In spite of Sebok's sophisticated treatment, the success rates evident from his totally inadequate - in terms of number of images - diagrams show the technique is not as effective as some of the parameter classifiers used in this Chapter.

JT have performed simulations of the effectiveness of their classification stratagy, but note that only simple poisson noise has been modelled, the conclusions concerning non-poisson noise sources discussed earlier suggest that the success rates derived from such a simulation are optimistic. The comparison performed between JT, Kron and Sebok techniques (JT) show that one or more of the classifier success rates are seriously overestimated. Although JT describe the results of the comparison with Kron's classifier as good, inspection of JT's Table VI reveals that below m_J=23 large discrepencies are evident between JT and Kron. The comparison itself is confined to the images with the best signal to noise that appear in both JT's and Kron's lists.

Suggestions that Kron and JT have overestimated their classification success rates have been graphically confirmed by Koo and Kron (1982). They have essentially performed a test of the same type as described in

this chapter - i.e. they examine the properties of faint images using more information than is available to the automated classifier: Koo and Kron use colour information whereas more detailed profile information is used here. Noting that galactic stars occupy a well defined locus in their three colour space Koo and Kron find disagreements of 25% between the number of colour classified, and profile classified stellar images. The colour properties of JT's stellar sample in the same field also suggest substantial contamination from extragalactic images. Koo and Kron's conclusions provide independent confirmation of the point stressed earlier in this chapter that it is not always possible to discriminate between stellar and non-stellar images within two magnitudes of the plate limit. A high degree of classification repeatibility between different plates - an observation on which JT lay considerable stress - without considering whether the technique is working satisfactorily does not guarantee a high degree of classification success.

5.8 Summary

A large catalogue of image classifications obtained from very deep plate material has been used to determine the success rates of COSMOS I.A.M. classification stratagies. The large intrinsic differences between objects of the same class as a function of magnitude leads to the use of several techniques to classify objects as a function of magnitude. Non-poisson noise sources in the plate emulsions together with the occurrence of merged images in the COMOS output set an upper limit to image classification success rates. The observed noise characteristics result in little improvement being gained by combining image classifications using different COSMOS I.A.M. parameters. Conclusions of this Chapter concerning the difficulty of eye and machine classification

within two magnitudes of plate limits are born out by image classification comparisons described in the literature, and the desirability of objective checking procedures during the development of image classification stratagies has been stressed. The magnitude limit for >90% integral probability of successful image classification on UKSTU J plates is brighter than previous determinations; corresponding to m_J=20.5 on the highest quality plates, with comparable limits relative to the plate limit for other plate types - i.e two to two and one half magnitudes brighter than the plate limits.

6 Correlation Function Estimates

6.1 Introduction

As described in Chapter 2 the correlation function technique has become the predominant form of galaxy clustering analysis, and is now finding application in virtually all astronomical clustering problems (e.g. Masson 1979, Osmer 1981, Shanks et al. 1980b, Bahcall and Soneira 1981, Sargent et al. 1980, Longair and Seldner 1979, Seldner and Peebles 1981). The theoretical basis for interpretation of correlation functions is well understood (Peebles 1973, Peebles 1980), and the ability to relate measurements of clustering from samples at different effective depths is one of the most attractive properties. This scaling behavior is a key feature of the samples analysed by the Princeton group (Groth and Peebles 1977; henceforth GP77), and they argue that the observed scaling lends strong support to their conclusions concerning the nature of galaxy clustering.

At the beginning of this project covariance function estimation software was written, and initial results derived from UKSTU and AAT plate galaxy samples. Large differences with some previous covariance estimates of the same samples were noticed. On further examination the methods of estimation were found to be somewhat different. These differences lead to a consideration of two main areas; (a) how to normalise the estimate of $W(\theta)$, and (b) how to treat data near the edges of the sample. The material of this Chapter together with some of the discussion of Chapter 2 has appeared as Hewett (1982).

The differences between the more recent covariance function results from machine data and those of the Princeton group have already been

discussed in Chapter 2. Examination of the estimation procedures used to derive the covariance estimates showed that very different techniques had been employed in some cases. Other reasons for differences between covariance estimates of samples of the same depth from different parts of the sky, and from samples of different depths are discussed in Chapter 9. In particular the affects of possible structure on scales of $50h^{-1}$Mpc should not be forgotten. A detailed investigation into the different types of estimator definition, and edge correction was undertaken to determine if the estimation techniques were responsible for any of the differences in the form of $W(\theta)$ evident between certain samples.

6.2 Estimation Of $W(\theta)$

Sharp (1979; henceforth Sharp) has considered some of the problems of edge correction when estimating $W(\theta)$ from a bounded sample. The consequences of differing estimator definitions as well as edge correction affects are emphasised in what follows, attempting to identify techniques as free as possible from systematic errors, rather than considering detailed error estimates as in Sharp.

6.2.1 Definition of the estimator $\hat{W}(\theta)$; the covariance function $W(\theta)$ is defined

$$\delta P = N^2 [1+W(\theta)] \delta\Omega_1 \delta\Omega_2 \qquad (6.1)$$

where δP is the joint probability of finding one object in solid angle element $\delta\Omega_2$ and another in solid angle element $\delta\Omega_1$ separated by an angle θ. N is the mean surface density of the sample.

When a sample contains angular positions of objects an estimate of $W(\theta)$ is

$$\hat{W}(\theta)= \frac{Np(\theta)}{N_c N \delta\Omega} -1 \qquad (6.2)$$

$Np(\theta)$ is the number of distinct pairs between $\theta \pm \Delta\theta/2$. N_c is the number of objects used as centers, N is the mean surface density of the sample and $\delta\Omega$ is the solid angle of the ring radius θ thickness $\Delta\theta$.

The equivalent Monte-Carlo estimator is

$$\hat{W}(\theta)= \frac{Np(\theta)}{Nr(\theta)} -1 \qquad (6.3)$$

where $Np(\theta)$ is the number of pairs in the sample with separations between $\theta \pm \Delta\theta/2$, and $Nr(\theta)$ the number of pairs for the same number of objects distributed randomly over an identical area. This method removes the need for separate calculation of edge corrections.

For a sample where counts Ni are in cells then one estimator (e.g. Shanks et al. 1980a; henceforth SFEM) is

$$\hat{W}(\theta)=\frac{\langle Ni*Nj\rangle}{N^2} -1 \qquad (6.4)$$

where the angular brackets $\langle\ \rangle$ denote averages over all cells with separations between $\theta \pm \Delta\theta/2$, and N is the mean surface density of the sample.

An important point is that the estimate of N in the definition of $W(\theta)$ - equation (6.1) - is derived from the total number of objects in the sample. If, as in practice $\hat{W}(\theta)$ is positive at small angular scales then it must become negative at some larger scale to satisfy the integral

constraint operating on the estimator, (see e.g. Peebles 1980; section 32). An example of the type of effect that can occur because of the constraint is illustrated in Fig. 6.1, where the data of Efstathiou and Eastwood (1981) is analysed using the estimator of equation (6.4). The simulation produces high contrast small scale structure with no large scale clustering present. Note how the estimator rapidly becomes negative after the initial small scale peak. The lower amplitude of $\hat{W}(\theta)$ in typical wide field 2D galaxy samples means the effect is smaller in practice, but for samples with linear scales at their effective depths comparable to the scale on which galaxies are strongly clustered - deep 4-meter plates and CCD images for example - this affect can be significant. For want of better terminology the estimators described above will be termed "direct".

SFEM in a major paper describing results from COSMOS measures of three UKSTU fields, proposed a modification to the direct estimator by employing a filtering technique (SFEM; section 3.3). This technique was applied to allow for large scale gradients in the galaxy distribution thought to be due to patchy galactic obscuration. A moving average filter is applied to the data, the estimator is calculated according to equation (6.4) and the count in the i-jth cell Nij is replaced by the value

$$N'ij=N*Nij/Ns \qquad (6.5)$$

where N is the mean count for the entire sample and Ns the mean count for the immediately surrounding area. The size of this surrounding area used to calculate Ns determines the filter scale length. SFEM used a filter length equal to half their sample size.

An "ensemble" estimator

$$\hat{W}(\theta)=\frac{\langle Ni*Nj\rangle}{\langle Ni\rangle\langle Nj\rangle} -1 \qquad (6.6)$$

has been employed by the Princeton group, (Peebles 1975) together with slight variations, (Peebles and Hauser 1974)

$$\hat{W}(\theta)= \frac{\langle Ni*Nj\rangle}{\langle (Ni+Nj)/2\rangle^2} -1 \qquad (6.7)$$

and by Dautcourt et al. (1978)

$$\hat{W}(\theta)=\frac{\langle (Ni-N)(Nj-N)\rangle}{N^2} \qquad (6.8)$$

Note the completely different approach to normalisation in the ensemble estimators: the normalisation is derived only from the regions of the sample used to calculate $\hat{W}(\theta)$ at each scale and the integral constraint does not apply to this type of estimator. Also shown in Fig. 6.1 is the Efstathiou and Eastwood data analysed using the estimator of equation (6.6). Note the different behaviour of the estimator at large scales where the spurious "anticlustering" produced by the direct estimator has been eliminated; only the small scale clustering is evident. In both cases the subsidiary maxima in the $\hat{W}(\theta)$ curves are due to the periodic nature of the simulation structure. For inhomogeneous or anisotropic samples - reflecting either large scale structure in the galaxy distribution or some external factors - it is possible that $\langle Ni\rangle \neq \langle Nj\rangle \neq N$ at one or more scales θ, and the direct and ensemble estimates of $W(\theta)$ will be different.

6.2.2 Edge correction; allowance must be made for the sample boundaries when estimating $W(\theta)$. Three techniques have been extensivly used

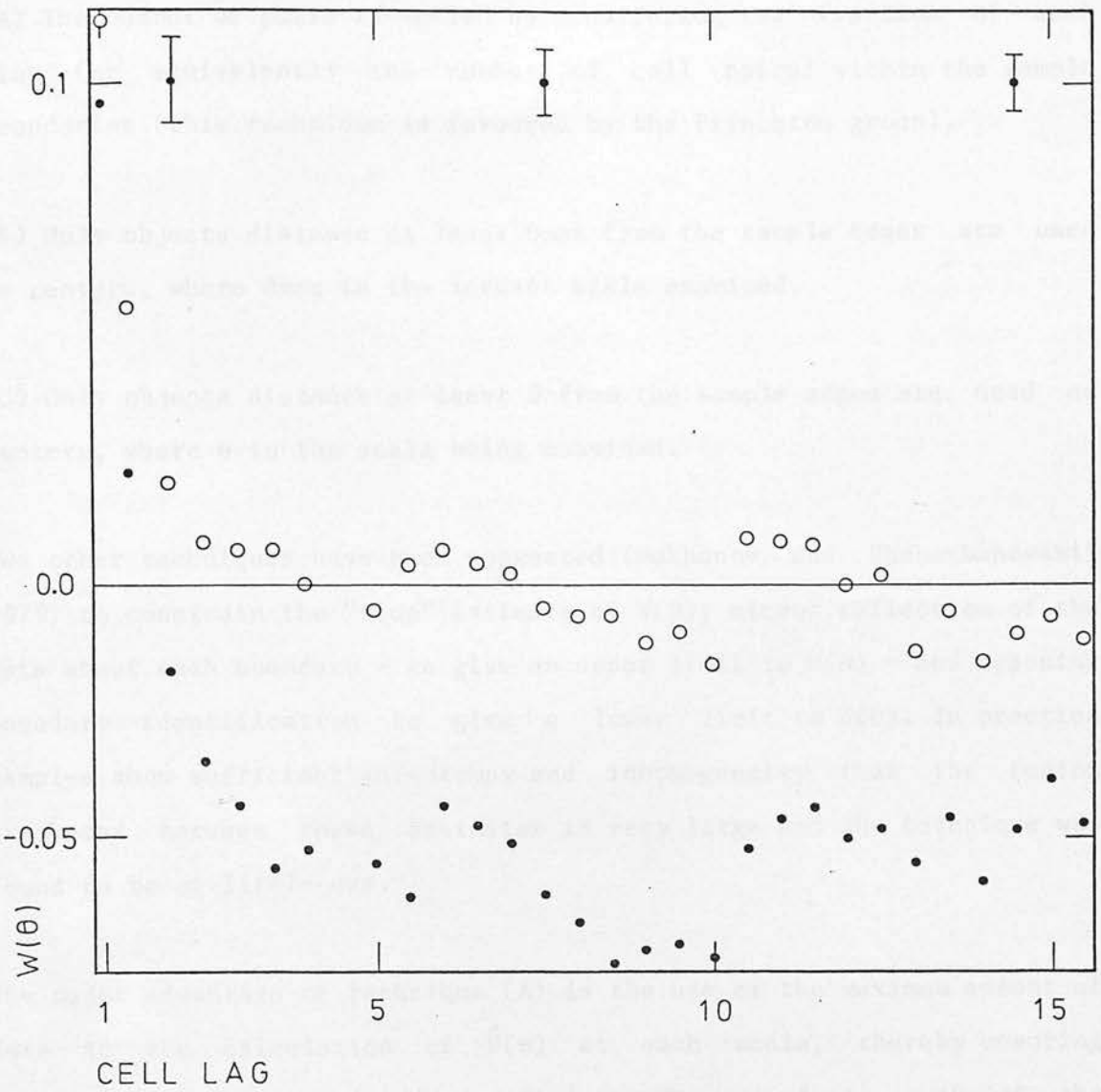


Fig. 6.1 Covariance function estimates from the Efstathiou and Eastwood simulation described in the text. Note the apparent anticlustering in the sample implied by the direct estimator (●) of equation(6.4). In contrast the ensemble estimator (O) of equation(6.6)is insensitive to the integral constraint which affects the direct estimator.

(A) The number of pairs is scaled by considering the fraction of each ring (or equivalently the number of cell pairs) within the sample boundaries (this technique is favoured by the Princeton group).

(B) Only objects distance at least θmax from the sample edges are used as centers, where θmax is the largest scale examined.

(C) Only objects distance at least θ from the sample edges are used as centers, where θ is the scale being examined.

Two other techniques have been suggested (Mukhanov and Shcherbanovskii 1979) to constrain the "true" estimate of $W(\theta)$; mirror reflection of the data about each boundary - to give an upper limit to $W(\theta)$ - and opposing boundary identification to give a lower limit to $W(\theta)$. In practice samples show sufficient anisotropy and inhomogeneity that the region bracketed between these estimates is very large and the technique was found to be of little use.

The major advantage of technique (A) is the use of the maximum amount of data in the calculation of $\hat{W}(\theta)$ at each scale, thereby ensuring statistical noise is reduced to a minimum. The use of as much of the data as possible to calculate $\hat{W}(\theta)$ also means that the method is least susceptible to systematic affects due to inhomogeneities and anisotropies in the data.

Methods (B) and (C) have been used in combination by Phillipps et al. (1978), MacGillivray and Dodd (1979) and Dodd and MacGillivray (1980). The fraction of the data used to calculate $\hat{W}(\theta)$ at each scale is much smaller than in method (A) so the noise is increased. Further, the methods are sensitive to inhomogeneities in the data as the fraction of

the data used to calculate $\hat{W}(\theta)$ changes with scale. In principle there should be no difference which method of edge correction is used providing the samples are homogeneous and isotropic, which ensures that any reasonable subsample of the data is a fair representation of the whole sample.

When considering samples as cell counts below, correction methods (B) and (C) are applied by considering only cell pairs where one or both are at least distance θ_{max} or θ respectively from the boundaries. The calculations on the galaxy samples in Section 6.5 are only carried out to a scale one quarter of the sample size for methods (B) and (C) because of the small fraction of the data contributing to $\hat{W}(\theta)$. Throughout this thesis the calculation of the noise associated with $\hat{W}(\theta)$ is made according to

$$(\delta\hat{W}(\theta))^2 = (1+\hat{W}(\theta))/Np \qquad (6.9)$$

where Np is the number of distinct pairs contributing to the calculation of $\hat{W}(\theta)$ at each scale. Sharp has shown that this is a slight overestimate in terms of "random statistical error". The 2σ error bars thus calculated shown in the figures are representative of those at neighbouring points which have been omitted for reasons of space. Data points have been joined up in the figures for clarity but no "interpolation" between data points is implied.

6.2.3 Cross-correlation of galaxy and random samples; Sharp suggested a method for investigating bias in the estimators of $W(\theta)$ when data is available as individual positions. The technique may be understood as follows: any homogeneous and isotropic sample of data should be unrelated to a set of points randomly distributed over the identical

sample area - the cross-correlation function between the data and random samples should be zero at all scales.

The cross-correlation function is defined

$$\delta P = N_1 N_2 (1 + W_{12}(\theta)) \delta\Omega_1 \delta\Omega_2 \qquad (6.10)$$

where δP is the joint probability of finding an object of type "1" in solid angle element $\delta\Omega_1$ and an object of type "2" in solid angle element $\delta\Omega_2$. The surface densities N_1 and N_2 of both samples enter into the equation and $W_{12}(\theta)$ is a measure of the cross-correlation of the samples, $W_{12}(\theta) = \theta$ for unrelated samples. Note the definition is symmetric $W_{12}(\theta) = W_{21}(\theta)$. Considering data as individual positions, then in practice Wgalaxy/random (Wg/r) will always be zero no matter how inhomogeneous the galaxy distribution, as the random sample is always randomly distributed about each galaxy - Wg/r is calculated by performing a sum over all the galaxies. In contrast choosing a random point as center and summing over all the random data to calculate Wrandom/galaxy (Wr/g) may give a nonzero result for an inhomogeneous or anisotropic galaxy distribution. This is equivalent to saying that the value of Wr/g is at least partially determined by the positions of the galaxies relative to the sample boundaries, e.g. when a large cluster is at the center of the field, (c.f. Fig. 5 of Sharp).

For counts tabulated in cells then Wg/r and Wr/g are no longer distinct. The estimators $\hat{W}r/g$ and $\hat{W}g/r$ are calculated by summing over both sets of objects so that $\hat{W}g/r$ is not necessarily equal to zero. It seems reasonable to correct the covariance function $\hat{W}(\theta)$ by subtracting the contributions of $\hat{W}r/g(\theta)$ and $\hat{W}g/r(\theta)$ from the original estimator. This is justified empirically using the data discussed below. A general

equation applicable to data both as individual positions and as cell counts is

$$\hat{W}corr(\theta)=\hat{W}(\theta)-\hat{W}g/r(\theta)-\hat{W}r/g(\theta). \qquad (6.11)$$

For samples as individual positions $\hat{W}g/r(\theta)=0$, and the equation is still valid. This equation is applied in Section 6.4 to obtain self consistent results from different estimator/edge correction combinations.

6.3 The Galaxy Samples

Three samples are used to demonstrate the affects of different estimation techniques

(1) The data of MacGillivray and Dodd (1980) consisting of 28872 objects. The area of the sample is 14.6 square degrees and full details of the sample are contained in MacGillivray and Dodd (1980). The covariance function for the sample published in Dodd and MacGillivray (1980) shows large discrepancies with the Princeton results.

(2) The Jagellonian sample analysed by Peebles (1975). The sample contains 12,145 galaxies in an area of 36 square degrees. The effective limiting magnitude of the sample is $m_B \sim 20.3$ (GP77).

(3) A south galactic pole (SGP) V sample which contains 5445 objects in an area of 4.8 * 4.6 degrees, with an area of 0.24 square degrees drilled out near bright images in the field, where the detection of galaxies by COSMOS is impaired. The sample is centered on the SGP and covers a somewhat larger area than that of SFEM's central field. The galaxy sample was obtained in parallel with the stellar sample in the

same field described by Reid and Gilmore (1982).

6.4 Results

Fig. 6.2 shows the covariance function estimates for the Jagellonian field using three different estimators - direct (equation (6.4)), filtered (equation (6.5)), and ensemble (equation (6.6)) - all using edge correction method (A). The filtered estimator was obtained using a filter length equal to half the sample diameter (i.e. the same scale relative to the sample size as employed by SFEM). Note the large differences evident between the ensemble, filtered and direct estimators shown in Fig. 6.2. For any particular sample the ensemble estimators of equations (6.6), (6.7) and (6.8) gave essentially identical results. In no case was a difference of greater than 0.002 found between any of the ensemble estimators of $W(\theta)$ for the three samples analysed. Equation (6.6) will be adopted as representative from now on. Similarly, for each sample analysed the direct estimators of equations (6.2), (6.3) and (6.4) gave identical results within the random statistical errors and for clarity the estimator of equation (6.4) is adopted as representative of direct estimators.

In Fig. 6.3, four different estimates of $W(\theta)$ from the Jagellonian sample using the filtered estimator of equation (6.5) with four filter scale lengths - 50%, 38%, 25% and 13% of the sample size - show substantial differences as expected. A filtering technique such as this should not be employed on data where many scales of clustering are present. The affects of a filter of this type with density fluctuations on several scales is complex - as shown by the Jagellonian results, a conclusion in agreement with recent comments on filtering by Fry and Seldner (1982). Filtered estimates of $W(\theta)$ will depend on where a single

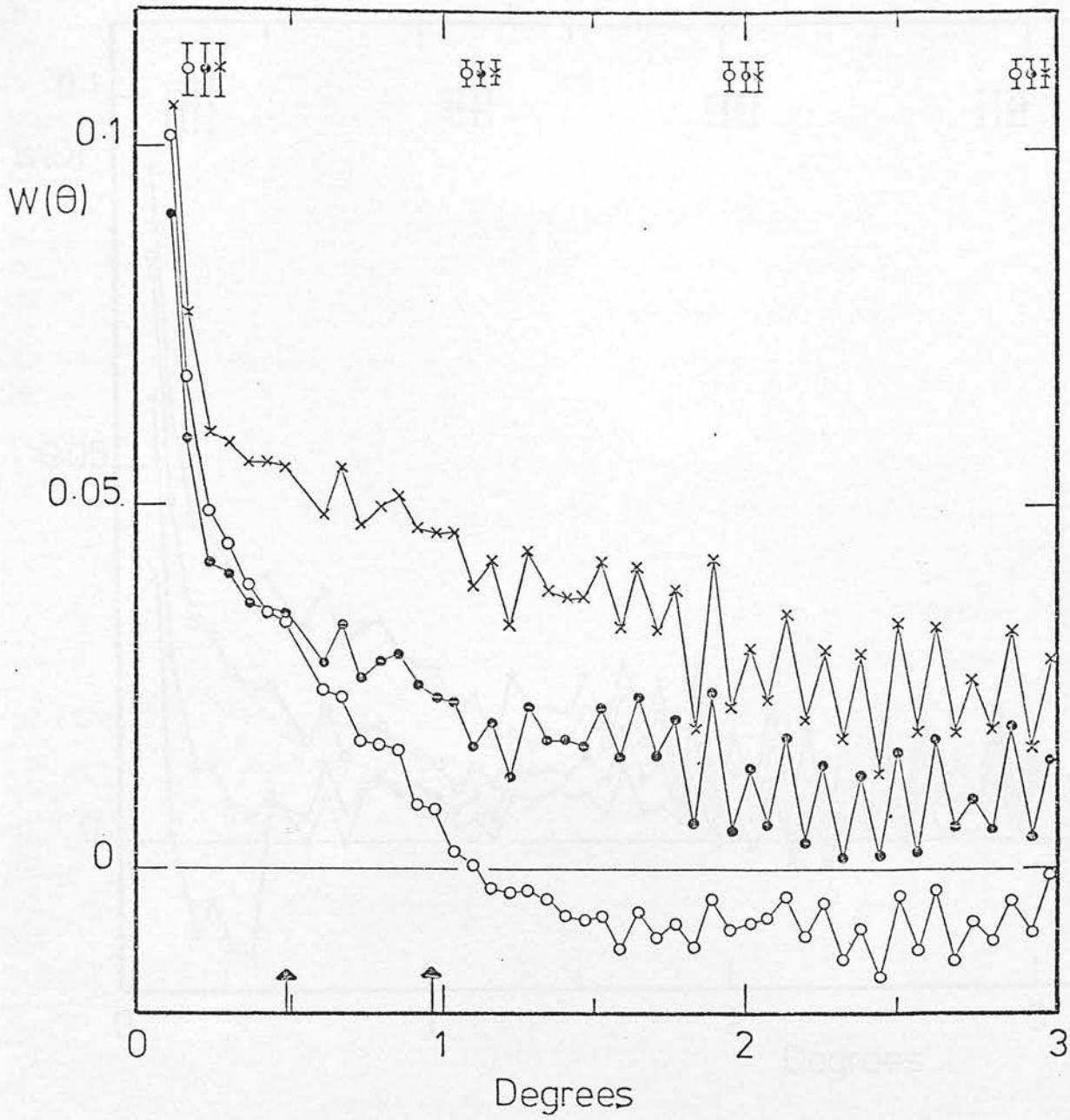


Fig. 6.2 Covariance function estimates for the Jagellonian field using the estimators of equations (6.4×), (6.5•) and (6.6○) all using edge correction method (A). Representative $\pm 2\sigma$ error bars are shown above the curves for each estimator. This convention for error bars is adopted throughout the chapter for clarity. Note the large differences between the three types of estimator.

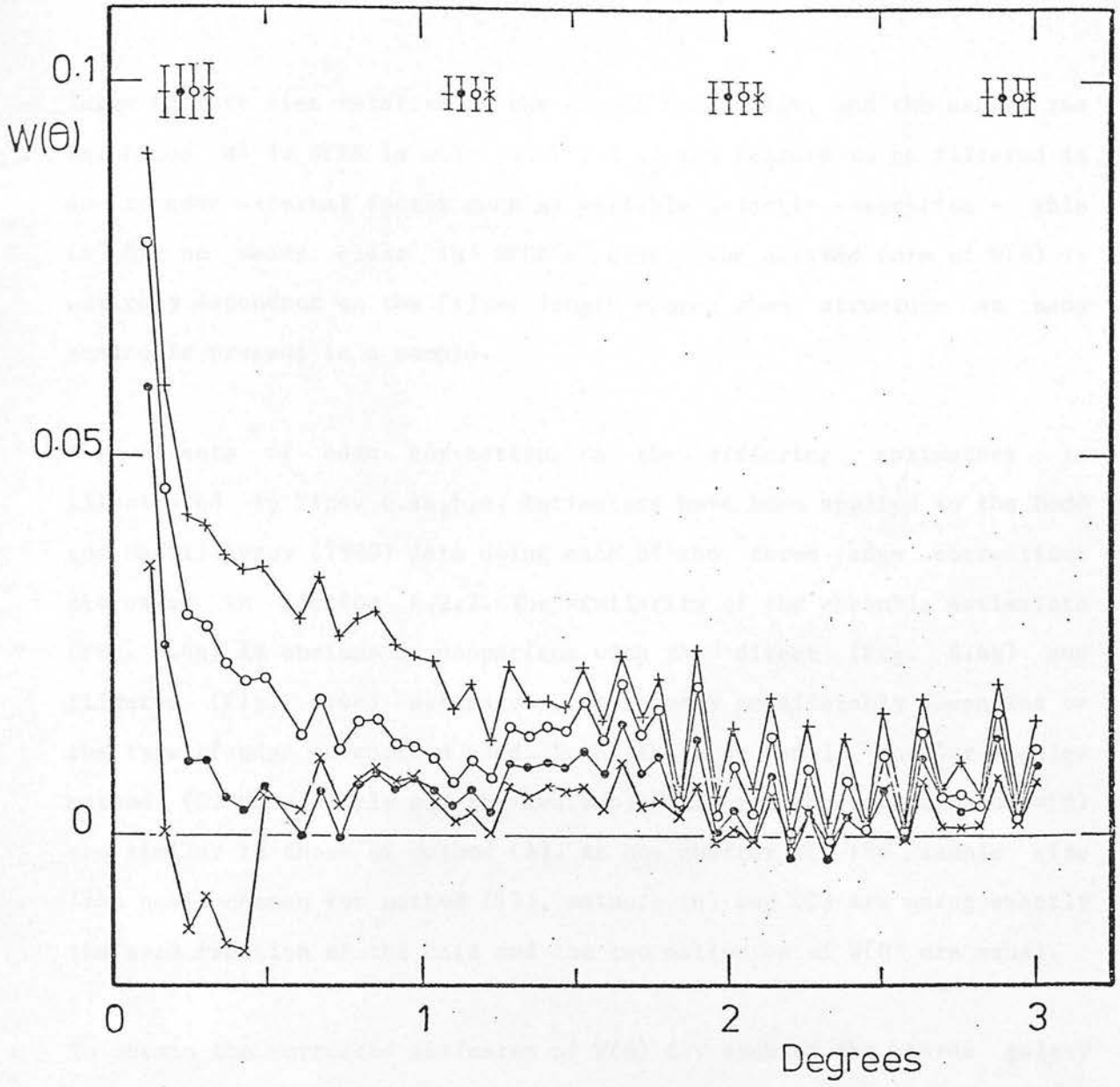


Fig. 6.3 Four estimates of the covariance function for the Jagellonian field using the filtered estimator of equation (6.5) employing different filter lengths. The filter scales as percentages of the sample diameter were 50%(+), 38%(o), 25%(•) and 13%(×). The results illustrate that the form of the estimators is completely dependent on the filter scale chosen. All the estimators were calculated using edge correction method (A).

large feature lies relative to the sample boundaries, and the use of the technique as in SFEM is only justified if the feature to be filtered is due to some external factor such as variable galactic absorption - this is by no means clear in SFEM's case. The derived form of $W(\theta)$ is entirely dependent on the filter length chosen when structure at many scales is present in a sample.

The affects of edge correction on the differing estimators is illustrated in Figs. 6.4a,b,c. Estimators have been applied to the Dodd and MacGillivray (1980) data using each of the three edge corrections discussed in Section 6.2.2. The similarity of the ensemble estimators (Fig. 6.4a) is obvious in comparison with the direct (Fig. 6.4b) and filtered (Fig. 6.4c) estimators, which vary considerably depending on the type of edge correction used. Note that at small angular scales method (C) uses nearly all the available data and the estimates of $W(\theta)$ are similar to those of method (A). At one quarter of the sample size (the scale chosen for method (B)), methods (B) and (C) are using exactly the same fraction of the data and the two estimates of $W(\theta)$ are equal.

To obtain the corrected estimates of $W(\theta)$ for each of the three galaxy samples under investigation five random data sets containing the same number of objects distributed over identical areas were generated. The galaxy samples were cross-correlated with each of the appropriate random samples. The average value of the cross-correlation calculated giving estimates of Wr/g and Wg/r for each sample. This procedure was followed for each of the estimator/edge correction combinations. The mean values of $\hat{W}r/g$ and $\hat{W}g/r$ for each of the equivalent $W(\theta)$ estimators in Fig. 6.4a,b,c are shown in Figs. 6.5a,b,c and 6.6a,b,c respectively. Values of $\hat{W}r/g$ and $\hat{W}g/r$ close to zero indicate that the correction to be applied to the original estimators is small. Conversly values of $\hat{W}r/g$

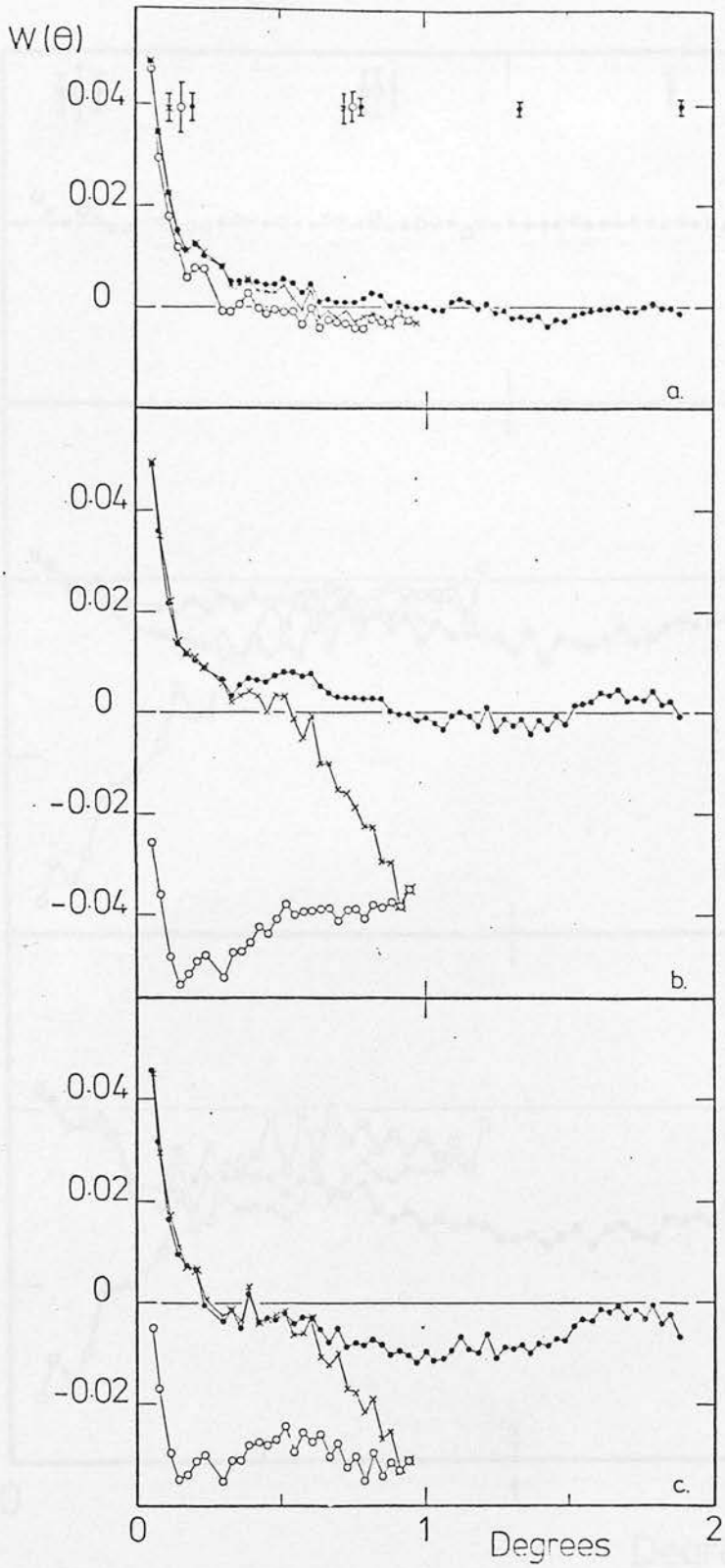


Fig. 6.4 (a) The ensemble estimators from the Dodd and MacGillivray (1980) field using edge correction methods A(●), B(○) and C(×). The symbol types for the edge correction methods are maintained throughout Figs. 6.4, 6.5, 6.6 and 6.7. Note the similarity beween the estimators despite the different methods of edge correction.

(b) Direct estimators for the Dodd and MacGillivray (1980) field with the three different edge corrections as in Fig. 6.4a. Large differences are present showing the estimator is strongly dependent on the type of edge correction employed

(c) Filtered estimators for the Dodd and MacGillivray (1980) field with the three different edge corrections as in Fig. 6.4a. As with the direct estimators shown in Fig. 6.4b the estimators show large variations.

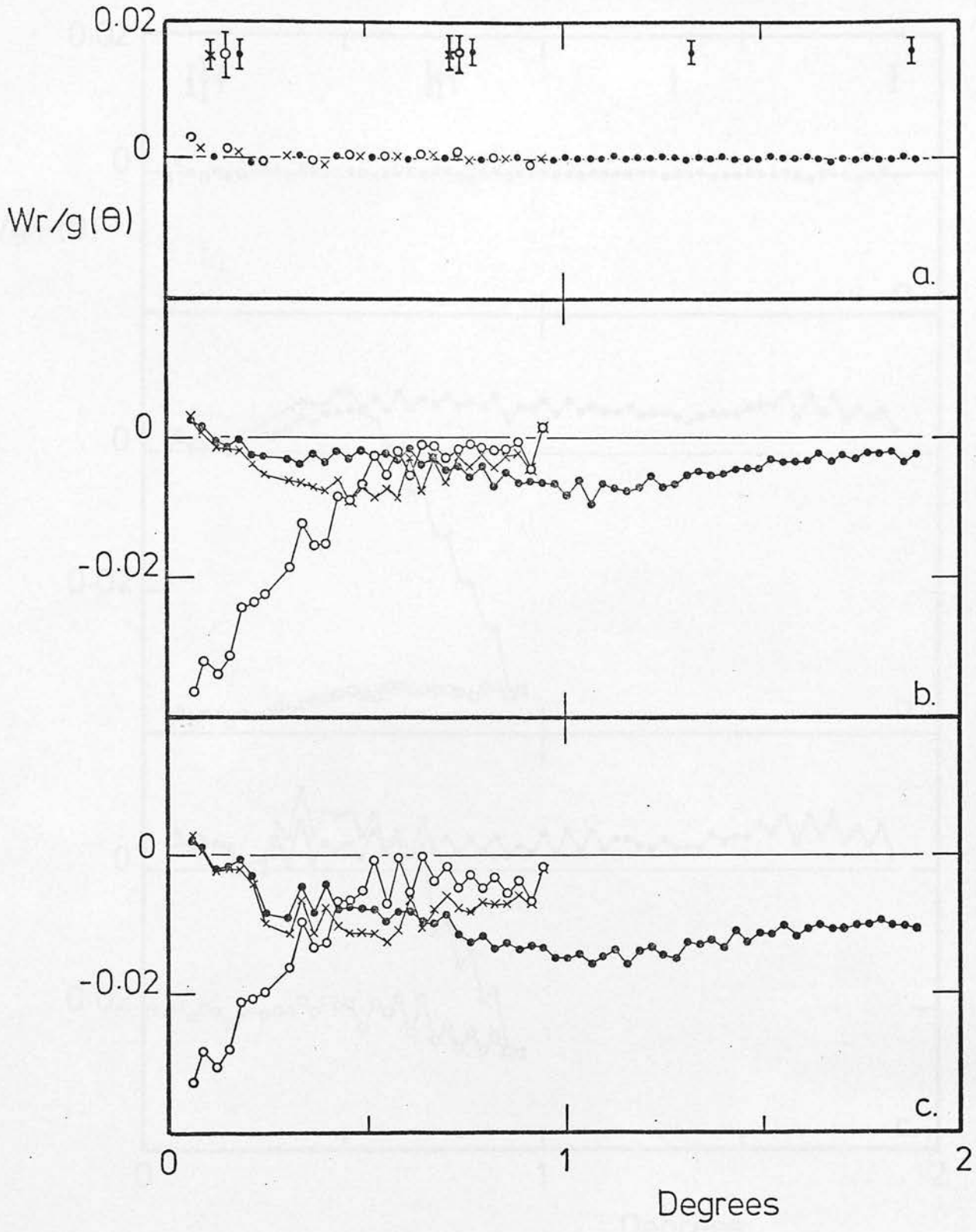


Fig. 6.5 (a) The mean cross-correlation corrections Wr/g for the ensemble estimators in Fig. 6.4a. The cross-correlations are equal to zero within the noise. Compare the virtually nonexistent corrections with those in Figs. 6.5b and 6.5c. Symbol types are as in Fig. 6.4 for each type of edge correction. As the values of Wr/g are so close to zero for the ensemble estimators only every third point for each method of edge correction has been plotted so that the individual data points are visible.

(b) Cross-correlation corrections Wr/g for the direct estimates of Fig. 6.4b. Significant deviations from the zero level are evident.

(c) Cross-correlation corrections Wr/g for the filtered estimates of Fig. 6.4c.

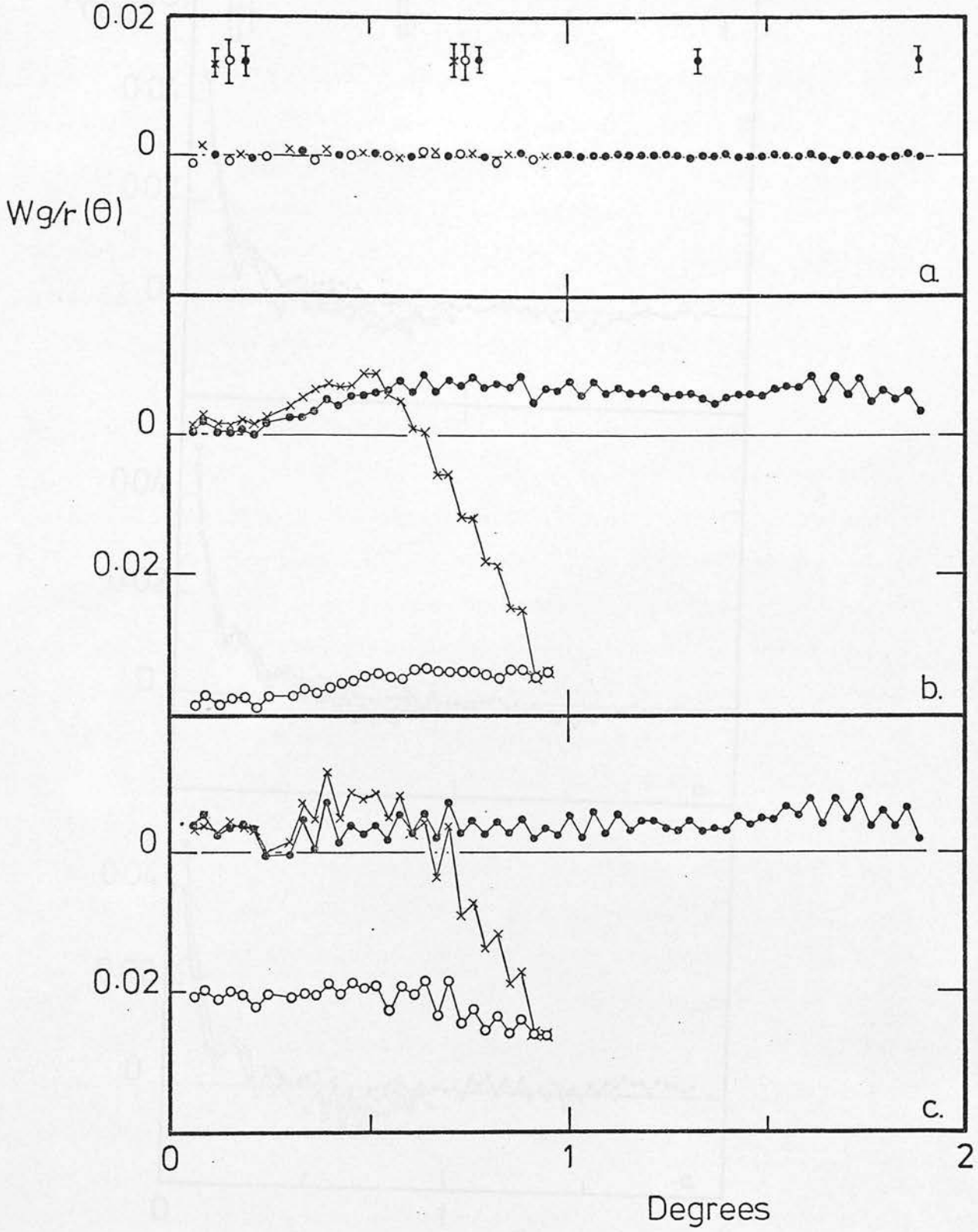


Fig. 6.6 (a) The mean cross-correlation corrections Wg/r for the ensemble estimators in Fig. 6.6a. Again as in Fig. 6.5a the cross-correlation corrections are equal to zero within the noise. As in Fig. 6.5a the cross-correlations for the ensemble estimators are so close to zero that only every third point has been plotted so that individual data points are visible.

(b) Cross-correlation corrections Wg/r for the direct estimates of Fig. 6.6b.

(c) Cross-correlation corrections Wg/r for the filtered estimates of Fig. 6.6c.

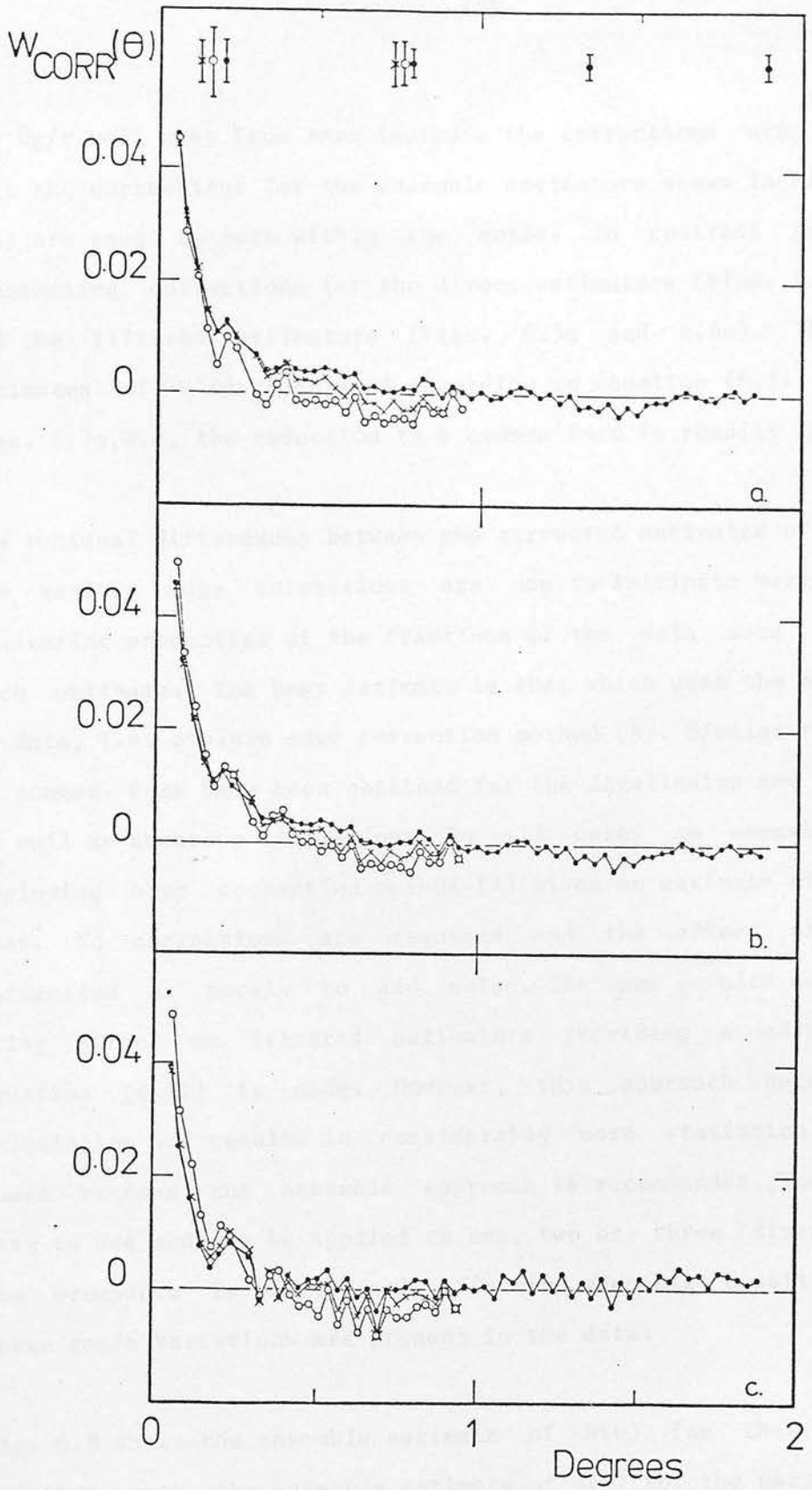


Fig. 6.7 (a) Corrected estimators for the ensemble estimators of the Dodd and MacGillivray (1980) field shown in Fig. 6.4a. The corrected curves are calculated according to equation(6.11) using the curves from Figs. 6.4a, 6.5a and 6.6a.

(b) Corrected estimators for the direct estimators of the Dodd and MacGillivray (1980) sample shown in Fig. 6.4b using the curves from Figs. 6.4b, 6.5b and 6.6b.

(c) Corrected estimators for the filtered estimators of the Dodd and MacGillivray (1980) sample shown in Fig. 6.4c using the curves from Figs. 6.4c, 6.5c and 6.6c.

and $\hat{W}_{g/r}$ well away from zero indicate the corrections are large. Note that the corrections for the ensemble estimators shown in Figs. 6.5a and 6.6a are equal to zero within the noise, in contrast to the large fluctuating corrections for the direct estimators (Figs. 6.5b and 6.6b) and the filtered estimators (Figs. 6.5c and 6.6c). The corrected estimates of $W(\theta)$ calculated according to equation (6.11) are shown in Figs. 6.7a,b,c, the reduction to a common form is readily apparent.

The residual differences between the corrected estimates of $W(\theta)$ using the various edge corrections are due to intrinsic variations in the clustering properties of the fractions of the data used to calculate each estimate. The best estimate is that which uses the maximum amount of data, i.e. employs edge correction method (A). Similar reductions to a common form have been obtained for the Jagellonian and SGP V samples as well as numerous simulations. In all cases an ensemble estimator employing edge correction method (A) gives an estimate of $W(\theta)$ free of bias. No corrections are required and the affect of making the correction is merely to add noise. The same results can be obtained using direct or filtered estimators providing a correction using equation (6.11) is made. However, this approach necessitates more calculation and results in considerably more statistical noise. For these reasons the ensemble approach is recommended. The procedure is easy to use and may be applied to one, two or three dimensional data. The procedure is particularly effective when high density contrasts or large scale variations are present in the data.

Fig. 6.8 shows the ensemble estimate of $W(\theta)$ for the SGP V sample together with the ensemble estimate of $W(\theta)$ for the parallel sample of stellar objects. This sample contains 15,145 stellar objects distributed over exactly the same area as the galaxy sample. The covariance function

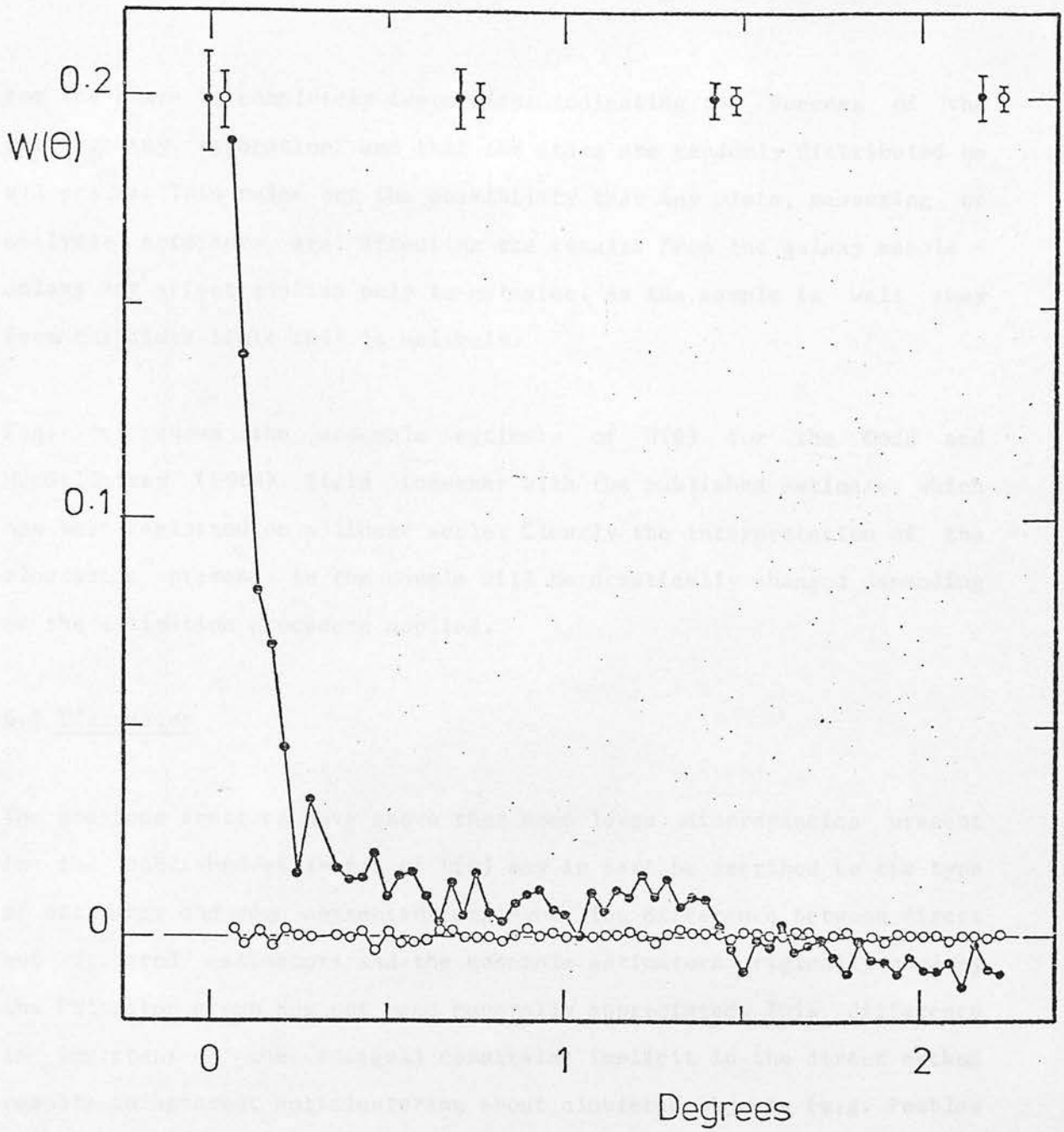


Fig. 6.8 The ensemble estimator from the SGP V galaxy sample (•) together with the ensemble estimator from the parallel stellar sample (o) showing that the low amplitude features in the galaxy covariance function are not due to any plate or machine measuring effects, unless the effect is confined to galaxies.

for the stars is completely featureless indicating the success of the star galaxy separation and that the stars are randomly distributed on all scales. This rules out the possibility that any plate, measuring or analysis artefacts are affecting the results from the galaxy sample - unless any affect applies only to galaxies. As the sample is well away from the plate limit this is unlikely.

Fig. 6.3 shows the ensemble estimate of $W(\theta)$ for the Dodd and MacGillivray (1980) field together with the published estimate, which has been replotted on a linear scale. Clearly the interpretation of the clustering present in the sample will be drastically changed depending on the estimation procedure applied.

6.5 Discussion

The previous sections have shown that some large discrepancies present in the published estimates of $W(\theta)$ may in part be ascribed to the type of estimator and edge correction employed. The difference between direct and filtered estimators and the ensemble estimators originally used by the Princeton group has not been generally appreciated. This difference is important - the integral constraint implicit in the direct method results in apparent anticlustering about clustered objects (e.g. Peebles 1980; sections 31 and 32), and explains why some samples show apparent anticlustering when strong clustering is present - e.g. Fig. 6.1. However the primary cause of the discrepancies evident in the data under discussion is the nature of the samples: the framework in which correlation functions are considered is one of homogeneous, isotropic and "fair" samples and in practice none of these three conditions are generally met. Large scale anisotropies are introduced into wide angle galaxy samples - e.g. the Zwicky and Lick counts - and most catalogues

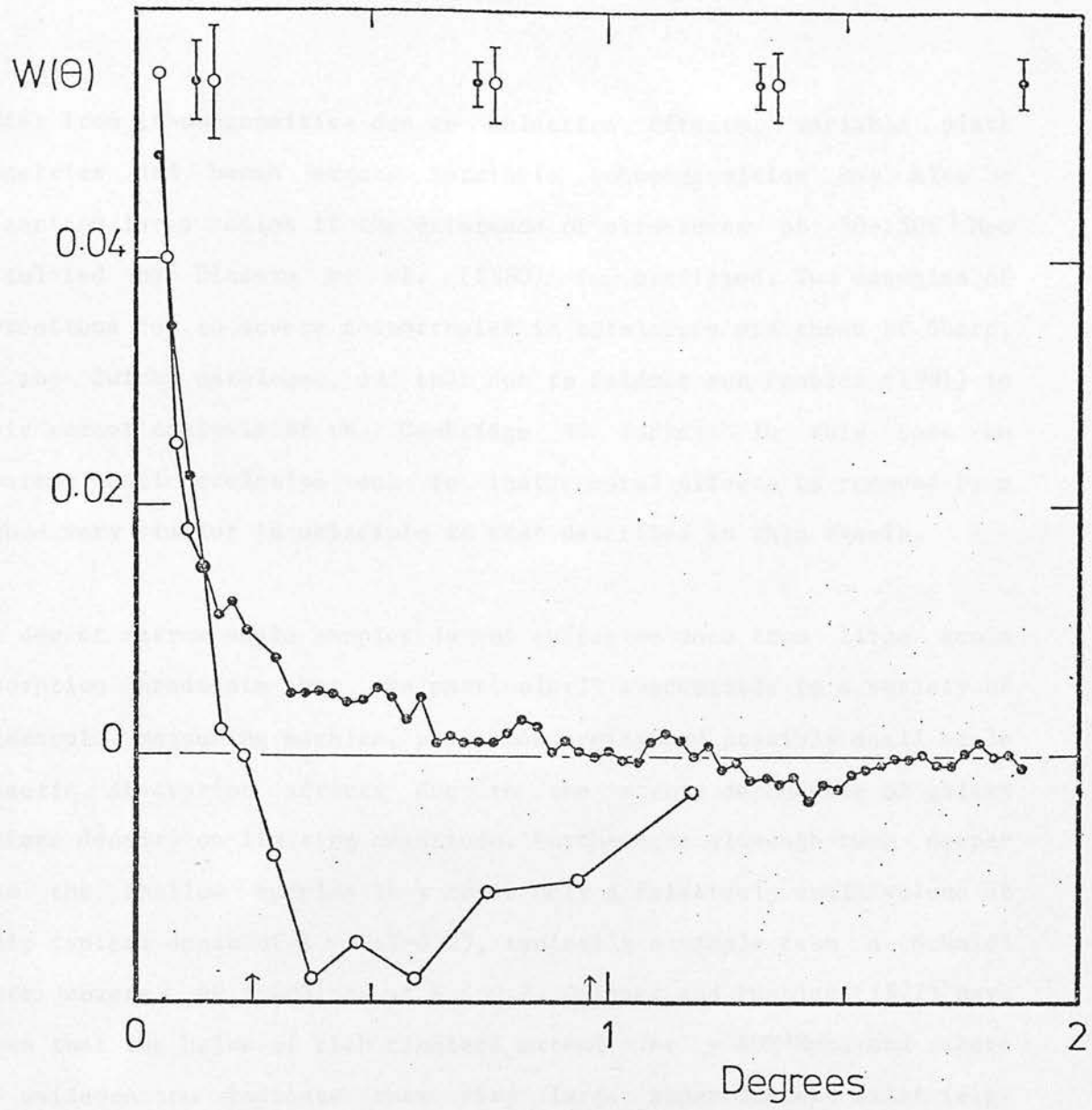


Fig. 6.9 The ensemble estimator from the Dodd and MacGillivray (1980) sample (●) together with the original estimator from the Dodd and MacGillivray (1980) paper (o).

suffer from inhomogeneities due to selection effects, variable plate properties and human error. Intrinsic inhomogeneities may also be present on large scales if the existence of structures of $50\text{-}150h^{-1}$ Mpc postulated by Einasto et al. (1980) is confirmed. Two examples of corrections due to severe anisotropies in catalogues are those of Sharp, to the Zwicky catalogue, and that due to Seldner and Peebles (1981) in their recent analysis of the Cambridge 4C Survey. In this case an apparent anticorrelation due to instrumental affects is removed by a method very similar in principle to that described in this thesis.

The deeper narrow angle samples do not suffer so much from large scale absorption gradients but are particularly susceptible to a variety of telescopic, measuring machine, plate uniformity and possibly small scale galactic absorption affects due to the strong dependence of galaxy surface density on limiting magnitude. Furthermore although much deeper than the shallow samples they cover only a relatively small volume at their typical depth of $Z \sim 0.2\text{-}0.25$, typically a sample from a Schmidt plate covers . $40 * 40h^{-1}$Mpc at $Z \sim 0.2$. Seldner and Peebles (1977) have shown that the halos of rich clusters extend over $\sim 40h^{-1}$Mpc, and there is evidence to indicate that very large superclusters exist (e.g. Gregory and Thompson 1978, Bahcall and Soneira 1983). Both of these considerations suggest that the samples analysed may generally not be fair and could be expected to show inhomogeneities and anisotropies on scales that would significantly affect the estimates of $W(\theta)$. The advantage of an ensemble estimators is the removal of affects due to the position of large scale features (whether due to the galaxy distribution or extrinsic causes) relative to the sample boundaries, allowing examination of the sample properties unaffected by the sample boundaries.

The Jagellonian data shows a strong feature at angular scale 1.0 degrees (i.e. a break from the power-law approximation at small scales) corresponding to a linear scale of $6h^{-1}$Mpc at the derived sample depth of $383h^{-1}$ Mpc (GP77). Beyond this scale the covariance function maintains small negative values until the limit set by the sample size, which corresponds to $18h^{-1}$ Mpc. No complex corrections were performed on the Jagellonian data to remove large scale inhomogeneities as was necessary for the Lick sample. This supports the view that the break is not an artifact of the correction procedures used by Groth and Peebles. The power-law form of the covariance function from the SGP V data is somewhat different from the GP77 result for the Lick survey. The results are discussed in Chapter 9, however, it is interesting that the results are in good agreement with those of SFEM at small scales - out to 0.25 degrees - but that major differences are evident at larger scales. Positive clustering is evident out to 1.4 degrees in the SGP sample, and this scale is comparable to the projected scale at which GP77 find a break from the power-law model for the Lick data, and to the scale of the feature in the Jagellonian field analysed here.

The confirmation of the existence of the break from the power-law form of the covariance function in both the Jagellonian and SGP V samples at scales in good agreement with that found by GP77 is strong evidence for the reality of the feature. Also of interest is the detection of weak anticlustering at large scales in both the Jagellonian and SGP V samples. The corrected estimate of $W(\theta)$ for the Zwicky catalogue from Sharp appears to show some evidence of anticlustering, this is evident in Fig. 1 of Bonometto and Sharp (1980), although it is not discussed. Examination of Fig. 2 in Sharp suggests that $\hat{W}(\theta)$ becomes negative at a scale ~ 15 degrees. The onset of anticlustering is difficult to measure precisely because of the low amplitude of the affect, however

qualitative scaling of the feature is certainly present: the approximate scales in the Zwicky, Jagellonian and SGP V samples at which negative values of $\hat{W}(\theta)$ occur are ~15 degrees, ~1.5 degrees and ~1.7 degrees corresponding to $13h^{-1}$Mpc, $10h^{-1}$Mpc and $11h^{-1}$Mpc respectively. The angular scales correspond to 19%, 25% and 39% of the sample diameter in each case. The onset of negative values of $\hat{W}(\theta)$ does not correspond to the same fraction of the sample size as might be expected if they were an artifact of the analysis technique. The interpretation of the negative values of $\hat{W}(\theta)$ is discussed in Chapter 9.

6.6 Conclusions

The discrepant estimates of $\hat{W}(\theta)$ for the Dodd and MacGillivray (1980) sample and almost certainly those of the MacGillivray and Dodd (1979) samples have been shown to be due to the analysis techniques employed. The existence of the power-law break point at scales entirely consistent with that found by GP77 in the Lick counts is established for the Jagellonian and a new SGP V sample, contrasting with evidence from recent machine measures that the position of the break point and the general form of the covariance function at scales beyond $3h^{-1}$Mpc are highly variable. A consequence of using ensemble estimators when calculating estimates of the covariance function - or correctly removing biases by cross-correlation with random data - is the detection of weak anticlustering in the Jagellonian and SGP V samples at angular scales corresponding to a linear size of order $15h^{-1}$Mpc. The use of an ensemble estimation technique in conjunction with edge correction method "A" of Section 6.2.2 to estimate $W(\theta)$ is strongly recommended for the calculation of correlation functions. When an ensemble technique can not be applied directly it is essential that corrections according to the prescription of Section 6.2.3 are made.

7 Data Reduction

7.1 Plates and COSMOS Measures

As described in earlier chapters the quantity of data was constrained by the quality of the plate material, and the availability of COSMOS measures free from large field effects. Finally, data for five areas of sky (covered by individual UKSTU plates) were accepted, consisting of three types; two areas with A grade IIaD emulsion V band plates of exceptionally high quality. These had the advantage that no detectable large scale background variations other than those expected from telescopic vignetting are present. Consequently, problems encountered with variable backgrounds on J and F band plates could be circumvented. The major disadvantage is the decrease in limiting magnitude compared with the J and F plates. The analysis of these plates was performed somewhat earlier in the project with many of the reduction techniques being developed using this data. Second, two areas had excellent glass copies of UKSTU A grade J survey plates, originally measured on COSMOS for other projects, the plates were measured again and made available for clustering investigation in mid-1982. Third, when it was not clear what caused the COSMOS field effects, and whether the problem was soluble in the time available, two unfiltered sky-limited IIIaJ plates were provided by UKSTU. Unfiltered plates are generally free from the large background variations caused by the Malin effect described in Chapter 4. The sky-background densities on these plates were too high (D=1.6) for successful measurement by COSMOS and glass copies were provided (background density D=0.8) by the ROE photolabs in early 1982. One of the copies was found to have a region of severe defocussing. After "drilling" of the apparently affected area, a strong anticorrelation between classified stellar and non-stellar images

remained, and the data had to be discarded. This experience confirms that of Corwin (1981): copy plates must be carefully checked before use for this type of project. Table 7.1 gives details of the positions and extents of the areas of sky covered by the plates used in the clustering investigation. Henceforth the designations SGP, MTF or the UKSTU field numbers - as listed in Table 7.1 - will be used to identify the areas.

TABLE 7.1

FIELD	RA	DEC	L	b	Total area (degrees)	Drilled area (degrees)	Areas drilled
MTF	22 03	-18 55	36	-51	23.1	0.5	14
SGP	00 53	-28 03	250	-89	22.0	0.7	6
297	01 44	-40 00	264	-73	23.1	0.5	10
350	00 24	-35 00	337	-81	21.7	0.4	7
352	01 12	-35 00	272	-81	21.0	0.7	14

For the SGP, MTF and F350 additional plate material with COSMOS measures was available, in the case of the SGP and MTF areas this material was of high enough quality to be of direct use in the subsequent reductions - as consistency checks for example. In the case of F350, some of the material had given problems earlier in the investigation, but was still of some use. In addition AAT plates of small portions of the SGP and MTF areas were measured. Table 7.2 summarises the primary plate material used in the investigation of the five fields, together with the COSMOS measures - details of UKSTU gradings are given in Cannon et al. (1978).

Note that three fields have multiple COSMOS measures of the primary plate made at different orientations, for the other fields single COSMOS measures of different plates were used. In all cases the COSMOS calibration and setting-up procedure was performed independently for

each measure. The measurement of different plates in the same area, and multiple measures of single plates is one of the primary reasons this investigation represents an improvement over earlier attempts to quantify galaxy clustering - the procedures described in chapters 3, 4 and 5 were carried out specifically to give more confidence in the data from the plate/machine combination than previously possible. This approach was carried over to the choice of plates - all of which were A grades.

Table 7.2

Field	Plate	Grade	COSMOS measures			
			Orientation	Threshold	Increment	Date
MTF	V5421	A	0	10%	16	14/7/82
	V2499	A	0	10%	16	26/5/82
SGP	V3475	AH	0	10%	8	7/2/81
	V6619	A	0	10%	8	19/2/81
297	J3593C	A2	0	7%	8	8/12/81
	J3593C		90	10%	8	6/6/82
350	UJ7154C	AI	0	10%	8	21/2/82
	UJ7154C		90	10%	8	24/2/82
352	J6124C	A	0	10%	8	7/12/81
	J6124C		90	10%	8	12/6/82

The single measures of different plates and multiple measures of single plates were paired together using the standard COSMOS software program written by B. McNally of IDPU. The original measures, together with the paired and unpaired images from the pairing program constituted the data for each field. Detailed examination of the data was performed to identify anomalous features due to differences between plates, or in the COSMOS measures. Table 7.3 gives the percentage success (relative to the number of images on the master plate) of the software pairing as a function of cosmos-magnitude for the five fields. Note the very high

success rates to cosmos-magnitude 0.0 on the J plates and -1.0 on the V plates.

Table 7.3

cosmos magnitude	J3593C	J6124C	UJ7154C	V5421/V2499	V3475/V2499
<-4	99%	99%	99%	99%	99%
-4 -3	99%	99%	99%	98%	99%
-3 -2	99%	98%	99%	95%	98%
-2 -1	99%	98%	99%	77%	83%
-1 0	97%	97%	98%	12%	21%
0 1	39%	85%	82%	-	2%
1 2	1%	57%	49%	-	-

(a) all percentages determined to better than one percent

(b) the rapid decline in pairing success for J3593C below cosmos-magnitude zero is due to the lower threshold cut of the master (7%) compared to the secondary (10%).

Table 7.4

	UJ7154C	J6124C	J3593C	V5421/V2499	V3475/V6619
cosmos magnitude	0.05	0.09	0.04	0.05	0.07
PSB	0.03	0.04	0.03	0.08	0.10
Log(area)	0.01	0.03	0.01	0.02	0.05
sky	1.2%	2.0%	1.7%	5.0%	8.0%
image density	2%	5%	6%	4%	2%

(a) quantities quoted are peak to peak between plates. Data calculated from averages over images in two cosmos-magnitude ranges; ≤ -2 and $-2 <$ cosmos-magnitude ≤ 0 for the J plates, ≤ -3 and $-3 <$ cosmos-magnitude ≤ -1 for the V data. Values are the maximum change in the two ranges.

(b) the large percentage sky changes for the V plates are due in large part to the different calibrations employed on each plate.

(c) the "image density" quantity refers to the image density change as a function of sky-intensity on each plate.

The paired data was analysed using the pairing test software written for the project, giving empirical determinations of the random and systematic errors for all the COSMOS I.A.M. parameters used in subsequent analysis - both as a function of cosmos-magnitude and position on the plate. Table 7.4 gives the peak changes in cosmos-magnitude, Log(area), sky-background intensity and peak surface brightness between the COSMOS measures of each field. The values refer to the peak variation between plates after the data from the pairing test program - calculated at a resolution of one square centimeter - had been block smoothed with a scale of three centimeters. The percentage change in image density over the range of sky-background intensity on each plate - from a linear least squares fit - are also given.

The use of multiple COSMOS measures fulfilled two essential functions; (a) determination of random errors due to COSMOS and plates, allowing accurate error modelling, and (b) identification of many types of field effects present on either the UKSTU plates or in the COSMOS measures.

Once the data had been verified in this fashion, object catalogues were obtained from each COSMOS plate measure, and from the paired measures, ensuring none of the individual plates or COSMOS measures gave anomalous results. The loss of images in the paired data was small; the percentage loss of images in the cosmos-magnitude range used for the object catalogues from measures of the same plate was less than 3% in all cases. For measures of two different V plates in the same field the percentage is larger, 2-3% is accounted for by noise images, and the fraction of astronomical images lost is still $\leq$9%. The main reason for the loss is the occurrence of two images appearing to be merged on one COSMOS measure but separated on a second, this results in one or two

images being removed from the paired data.

7.2 Preliminary Data Reduction

COSMOS ellipse plots of the areas surrounding the brightest twentyfive images of each measure were examined to determine if significant numbers of spurious images were present; around bright objects COSMOS breaks the low surface brightness portions of galaxies or halation surrounding bright stars into many "galaxy" images. The severity of the effect depends on the type of background following used in the measure. It is essential to remove these noise images, as they form small high density "galaxy clusters" in the data (see Corwin 1981; Fig. 4.9 for example) - and can contribute significantly to clustering statistics, especially on small scales. On most plates a few very bright objects significantly bias the background in their vicinity by > 10%, making image parameters unreliable, and any region where this occurred was identified, regardless of whether many noise images were present. Having identified the positions and extents of the regions effected by noise images or background contamination, these areas were removed; rectangular areas covering the effected regions were "drilled" out of the data. Further discussion, including the frequency of noise images, refers to the drilled data. Any other areas where the pairing comparison indicated localised systematic changes in COSMOS parameters (presence of stepwedge nearby), satellite trails or other non-astronomical noise - were also removed. The fraction of the measured area lost is small - Table 7.1 - and a considerable improvement over earlier data - e.g. SFEM Figs. 2 and 3. The number and extent of the regions requiring "drilling" depends critically on the scale-length and threshold adopted in the COSMOS background following. Whenever the sky-background is significantly biased by the presence of a bright image the data should be discarded.

However, the very large scale-length adopted by COSMOS in 1982 (±1cm) as the default for background following results in regions where the sky-intensity is only a few percent brighter producing large numbers of spurious images. This is highly unsatisfactory and the measures of plates V5421 and V2499 required a large number of drilled regions due to this problem. The detailed data presented in this and subsequent Chapters refer to the zero degree orientation COSMOS measure of F350, the ninety degree orientation measures of F297 and F352, and to the measures of V3475 and V5421 for the SGP and MTF fields respectively. The data for the paired measures, second COSMOS measures or other plates are in no way significantly different. This is to be expected from the completeness data and lack of detectable field effects - Tables 7.3 and 7.4. Table 7.5 gives the percentage of noise contamination and completeness limits for the deepest samples determined by eye inspection, where possible completeness limits between different plates are given. Calculations are performed in terms of image number density - see Chapter 9. The noise frequency in the "bright" and "intermediate" galaxy samples is negligible and less than 1.5% in all cases.

7.3 Image Classification

To improve the signal to noise of the observed clustering, and to ascertain whether variable galactic extinction is important in the fields, a separation between stellar and non-stellar images is desirable. Problems with classification of bright images in COSMOS I.A.M. data have already been described - Chapter 5. To obviate the difficulties in this region the brightest 700 images on each field were classified by eye. This had the additional benifit of achieving an overall familiarisation with the properties of each field - positions of

obvious galaxy clusters for instance. The faint magnitude limit of the eye classified images was $m_J \sim 15$ corresponding to a cosmos-magnitude limit of approximately -6.5. Images fainter than this were classified using techniques described in Chapter 5.

Table 7.5

	UJ7154C	J6124C	J3593C	V5421	V3475
cosmos magnitude limit	0.0	-0.7	0.2	-1.0	-1.0
image density	4650	5170	4540	2630	1400
completeness control	J2633	J3570	J3587	V2499	V6619
noise	4%	4%	3%	2.5%	2.5%
Completeness	88%	89%	90%	92%	91%
	91%	93%	93%	94%	93%

(a) all percentages determined to $\pm$1%. Number densities to better than $\pm$25.

(b) first "Completeness" column refers to the data with noise, second to the data without noise. Approximately 3% of the incompletness is due to random errors in the COSMOS measures.

The small number of plates finally reduced allowed considerable attention to be paid to the separation, and determination of the classifier success rates. The star-galaxy separation for all the fields was performed by the interactive definition of separation lines in cosmos parameter spaces using the information from the eye classifications. Three parameter spaces were employed: cosmos-magnitude vs. Occupation Index, Log(area) and Peak Surface Brightness. The Occupation Index criteria was not necessary for the V plate data where saturation effects and diffraction spikes are not severe. The application of the automated classifier described in Chapter 5 gave

virtually indentical results in terms of classifier success.

Table 7.6

Field	Sample	Galaxy density	Stellar density	Galaxies lost	Galaxy fraction
SGP	B	220+20	720+20	1+1%	.90+1%
	I	510∓15	915∓15	1∓1%	.72∓3%
MTF	B	265+30	1580+30	2+1%	.57+4%
	I	590∓20	1960∓20	2∓1%	.46∓4%
297	B	240+20	680+20	3+1%	.82+2%
	I	925∓40	920∓40	19∓2%	.90∓1%
	F	2590∓150	1950∓150	7∓1%	.70∓5%
350	B	300+20	770+20	1+1%	.78+2%
	I	790∓40	1010∓40	9∓2%	.85∓1%
	F	2750∓150	1900∓150	3∓1%	.73∓5%
352	B	260+20	590+20	1+1%	.78+2%
	I	900∓40	960∓40	6∓1%	.83∓1%
	F	3070∓200	2100∓200	2∓1%	.72∓5%

(a) surface densities are per square degree.

(b) the designations B, I and F stand for "bright", "intermediate" and "faint" samples.

(c) the fraction of galaxies lost is expressed as a percentage of the total number of galaxies. The Galaxy fraction is the fraction of the "galaxy" sample that are actually galaxies.

Eye classification of 2000-3000 images in each field was performed; multiple plates were used, and in the SGP and MTF fields deep AAT plates were employed. Several areas were examined in each field spanning the measured area of the plates to ensure that the classification success rates did not vary as a function of plate position. The resulting eye classifications are - for the reasons described in Chapter 5 - far more reliable than those previously used for galaxy clustering investigations. The classification discrimination lines were chosen to include virtually all galaxies, with only "definite stars" being excluded. Fainter than cosmos-magnitude -2.0 virtually all images are

assigned to the "galaxy" class. The makeup of the "galaxy" samples are given for each of the fields in Table 7.6. Note the very high proportion of galaxies included in the "galaxy" sample. With the exception of the F297 intermediate sample, there is no significant misclassification of galaxies using this approach. This procedure prevents the loss of significant numbers of the most compact galaxies at the expense of more stellar contamination.

The data from the paired, and unpaired COSMOS measures of the five fields with image classification and "drilling" performed as described in this chapter form the object catalogues analysed in Chapter 9.

8 Variable Galactic Extinction

8.1 Introduction

The amplitude and nature of extinction fluctuations (caused by interstellar dust grains) in the plane of the galaxy has remained a controversial subject for the field of galaxy clustering (e.g. de Vaucouleurs 1971, Holmberg 1974, Burstein and Heiles 1978). If the clustering intrinsic to the galaxy distribution is to be determined, the amplitude and scale of any spurious clustering induced by variable extinction along different lines of sight through the galaxy must be quantified. The existence of variable extinction results in the effective limiting magnitude of galaxy counts varying with position on the sky: the implication of such variations for the detection of low amplitude density enhancements was stressed in Chapter 1. To ascertain the affect of variable extinction on the galaxy samples analysed in Chapter 9 a method to determine limitations on the amplitude of fluctuations at small angular scales in the extinction layer was developed using the companion stellar distributions to the galaxy samples of each field.

8.2 Properties of the Extinction Layer

The simplest extinction model assumes that dust is uniformly distributed in a plane parallel layer coincident with the galactic plane - this leads to the well known cosecant variation of extinction as a function of galactic latitude - an approximation regularly applied in many clustering studies to remove large scale gradients in the distribution of extragalactic objects due to the global extinction variations (e.g. Hauser and Peebles 1973).

Little further observational data relating to possible smaller scale variations was obtained before the mid-1970s. Earlier, attempts were made to determine the properties of the extinction layer using galaxy counts themselves (e.g. Shane and Wirtanen 1967, Holmberg 1974). The somewhat circular arguments implicit in much of this work has caused considerable confusion over the mean extinction at the galactic poles, and the amplitudes of fluctuations in the extinction layer. Fesenko (1978) and Holmberg (1974) for example ascribe variations in the galaxy counts on all but the smallest angular scales to variable extinction, suggesting large fluctuations in the extinction layer with little intrinsic galaxy clustering present on scales greater than a few megaparsecs. The belief that significant fluctuations occur in the extinction layer at high galactic latitudes is still held by several groups (e.g. SFEM, Tyson and Jarvis 1980).

The observational evidence relating to the properties of the extinction layer consists of four broad types: (a) large scale mapping of HI emission, (b) high precision intermediate band photometric studies of early type stars, (c) interstellar polarisation observations, and (d) HI emission and absorption studies toward specific objects. The observational situation is comprehensively reviewed by Burstein and Heiles (1978, 1982; henceforth BH82) and a detailed review is not presented here. BH82 conclude that deductions concerning extinction fluctuations at high galactic latitudes using galaxy counts are unreliable; the major fallacy in this approach is the assumption that all large scale structure observed in the galaxy distribution is due to variable extinction - this neglect of the intrinsic large scale structure results in gross overestimates of the amplitude of fluctuations. The redshift studies discussed in Chapter 2 clearly

demonstrate the existence of large scale ($>10h^{-1}$ Mpc) structure in the galaxy distribution, the amplitude of which is an order of magnitude or more greater than can be explained by extinction fluctuations. The picture emerging from other observations of the extinction layer is not clear cut; there is disagreement over the absolute values of HI column densities at high latitudes for instance (Kalberla 1980a, 1980b, Heiles et al. 1981). However the debate is not of direct relevance as proponents of high HI column densities find a large fraction of the gas to be uniformly distributed (Heiles 1980) - i.e. not clumpy.

The angular resolution of current data is confined to scales above two degrees: small scale structure is not resolved. The majority of the observations suggest a mean extinction $E_{b-v} \leq 0.025$ for galactic latitudes $b \geq 50^{\circ}$, with variations of a factor two evident on scales of ten degrees (e.g. Hilditch et al. 1976, BH82). A minority of workers find mean values of $E_{b-v} \geq 0.05$, but BH82 cast doubt on the validity of much of this evidence. The contrast between evidence for $E_{b-v} \leq 0.025$ from galactic tracers and the galaxy number count data giving $E_{b-v} \geq 0.12$ is remarkable.

One parsec subtends an angle of 0.6 degrees at a distance of one hundred parsecs, and if the dimensions of clouds found in the plane of the galaxy (e.g. Knude 1981) are similar at high latitudes then significant fluctuations can be expected on scales of several degrees. For the galaxy samples of Chapter 9 the range of scales 0.1 - 3 degrees is relevant; precisely the range where fluctuations are expected for reasonable cloud sizes and for which there is little data. This situation is unlikely to improve due to the lack of suitable objects for reddening determination with a high enough surface density to give improved resolution. It is important to ensure the results of Chapter 9

are not significantly affected by variable extinction.

8.3 Scaling Effects of Galactic Extinction

It is frequently assumed that apparent clustering induced by patchy galactic extinction can be recognised by its occurrence at fixed angular scale, independent of the galaxy sample depth. This is correct in principle for the simple case of one dominant scale of extinction variations and idealised galaxy samples. However there are complications. A single cloud size with clouds located at a fixed distance is unrealistic; the distribution of extinction with z (distance from galactic plane) follows an exponential or gaussian law (e.g. Dickey et al. 1981) with scale-height 80-120pc. There is a huge range in apparent scale of extinction fluctuations from features extending many tens of degrees (e.g. Gould's belt) to fractions of a degree - individual small clouds: the size spectrum of features in the galactic interstellar medium (ISM) results in apparent extinction variations over scales of 0.25-30 degrees. In a galaxy sample the detection of extinction effects requires reasonable signal to noise relative to the fluctuations intrinsic to the galaxies themselves; either a large number of galaxies must be affected by each cloud or the survey area must be large - to achieve high S/N by averaging over the small affects of many clouds. For bright nearby samples (e.g. the Zwicky catalogue) the intrinsic clustering is strong on scales of a few degrees, and small scale extinction variations will be very difficult to detect. At larger scales (>10 degrees) the intrinsic clustering is weak and many galaxies will be affected by extinction gradients on these scales. At fainter magnitudes the amplitude and scale of intrinsic galaxy clustering decreases rapidly due to projection - e.g. a factor 80 between 15th magnitude Zwicky sample and a 22nd magnitude UKSTU sample. So that

extinction fluctuations at fixed angular scale have much greater influence on fainter samples. A cloud subtending one degree affects one galaxy on average in the Zwicky catalogue and 2500 in a 22nd magnitude sample - the S/N increase with magnitude is enormous. Further, entirely new small scale fluctuations that have no detectable affects on shallow wide angle samples may become important at faint magnitudes. Conversly very large scale gradients will not significantly influence the deep samples, which to date have been confined to small areas of sky. The presence of extinction fluctuations on many scales, variable angular sizes of galaxy samples to different depths, and strong dependence of effective S/N on the intrinsic galaxy clustering at a particular scale, means that scaling behaviour due to variable extinction is more complex than sometimes appreciated.

8.4 Detection of Variable Extinction Using Stellar Samples

A by-product of the galaxy samples described in Chapter 9 are samples of stellar objects with low contamination from galaxies and extragalactic objects - Chapters 5 and 7. The scale-height of gas and dust in the galaxy is small; observations give values in the range 80-120pc (e.g. Baker and Burton 1975) - most extinction is very close (>65% within 100pc of the sun). In contrast, the z distribution of galactic stars is dominated by components with large scale-heights (Reid and Gilmore 1982; henceforth RG82). RG82 find the distribution of stars at the SGP is well modelled by two components; one consisting of 90% of the galaxy's projected stellar surface density with scale-height 300pc, and a second composed of 10% of the stellar surface density with scale-height 1500pc: more than 75% of the stars observed toward the SGP will be affected by greater than 65% of the total line of sight extinction - figures for a stellar sample limited at m_J =20, ISM scale-height 100pc.

Stellar samples of 1000 objects per square degree (m_J 20) can be obtained from UKSTU J plates, and 750 per square degree from V plates - the potential for detecting extinction fluctuations via the apparent clustering of stars is good. However, the form of the stellar number counts is not so favourable at faint magnitudes; the slope of the integral number counts is approximately 0.2*m at 20th magnitude in B (e.g. Allen 1973). The possibility of using the stellar samples to place limitations on the fluctuations in the extinction was investigated further.

Key variables for extinction fluctuations are the mean exinction over the line of sight, and the extinction and size of an individual cloud. Assuming a poisson distribution of clouds of fixed size, and extinction per cloud, then the dependence of fluctuations on these quantities is readily derived. If "k" is the mean number of clouds per line of sight, and Ec the extinction per cloud: the average extinction is <E>=k*Ec, and using the poisson distribution the fluctuations in the extinction depend on sqrt(k)*Ec - i.e. the amplitude of fluctuations is directly proportional to the extinction per cloud, and to the square-root of the mean extinction. The size of the clouds determines the scale of the fluctuations. To investigate in detail the affects of variable extinction on galaxy and stellar samples a simple simulation program was developed.

8.5 Method of Simulation

A conical region of the galaxy was modelled; vertex at the observer, angle determined by the width of the field to be modelled and the angular extent of the largest clouds. Cloud positions were generated randomly with a suitable bias to reproduce the desired z distribution.

The extinction from each cloud was calculated, and an array of extinction values (normal to the line of sight) calculated by summing all the cloud contributions along each line of sight. Resolution in the z direction was provided by calculating extinction arrays for every twentfive parsecs in z. The angular resolution of the arrays was one to four arcminutes.

Stars were generated; magnitudes and positions drawn from the adopted luminosity and z distributions, the apparent magnitude of each star was modified by the line of sight extinction to the star, and accepted in the sample if the apparent magnitude was less than the chosen sample limit. Stellar samples to any limiting magnitude, modified by a chosen extinction model could be generated. Galaxy samples were produced in a similar fashion except that each galaxy is affected by the total line of sight extinction. The parameters for the galaxy distribution were those of the "standard model" of Table 9.1. Reasonable variations in the parameters do not change the conclusions of this Chapter - the amplitude of the effect on the galaxies is primarily determined by the slope of the number counts which are very insensitive to the selection function parameters. Two questions were addressed: which extinction models severely modified the apparent clustering of faint galaxy samples, and could these models be ruled out by considering the amplitude of the stellar clustering.

The main input parameters to the simulation are (a) overall distribution of stars; taken from RG82, with 90% of the surface density described by an exponential fall off from the plane with scale-height 300pc, the second component (10% of stars) has an exponential scale-height of 1500pc - the number of young disc stars at the poles is negligible and can be ignored. The model parameters are insensitive to changes in the

scale-height of the more diffuse component - increasing the scale-height of the major component improves the upper limits on extinction derived below, conversly a reduction in scale-height results in the limits becoming weaker. For changes in scale height of $\pm$ 50pc the percentage change in the amplitude of the stellar covariance function (Section 8.7) are ($\pm$ 8%). As RG82 discuss, there is little evidence to indicate that the scale-height of the majority of galactic stars is significantly less than 250pc. The luminosity function for stars was taken from Wielen (1974), the range of magnitudes modelled was $0< m_V <16$, RG82 find evidence for a cutoff at the bright ($m_V \leq 5$) end of the luminosity function for the extended component - this was incorporated in the simulation. The use of the RG82 or Luyten (1968) luminosity functions does not affect the model conclusions. This can be readily seen because all determinations are similar at intermediate magnitudes; the number of stars at bright magnitudes ($m_V \leq 2$) is very small (<1% of the space density) and the volume in which very low luminosity ($m_V \geq 16$) stars can be seen is small (distance limit for detection of $m_V = 16$ star at 20th magnitude is 60pc), so that the model is insensitive to the extremes of the luminosity function.

(b) The distribution and extinction properties of galactic dust are reviewed by Savage and Mathis (1979) and a full review is not presented here. The overall distribution of dust and gas is not exactly determined, but there is reasonable agreement between workers: gaussian and exponential models for the z distribution with model parameters equivalent to exponential scale-heights of 80-120pc are favoured (e.g. Dickey et al. 1981). The simulation is not sensitive to the exact z distribution but if the scale-height has been significantly underestimated the limits derived in section 8.7 will be seriously affected; changing the dust scale-height to 150pc results in a change of

20% in the amplitude of the stellar covariance function. There has been much discussion concerning the values of the ratios R=Av/Eb-v (the ratio of absorption to selective extinction), and the dust to gas ratio. However, recent evidence suggests that both ratios are more nearly constant than earlier observations indicated (Savage and Mathis 1979, Strong and Lebrun 1982). If these ratios do vary at high galactic latitudes this will increase the fluctuations and strengthen the conclusions of Section 8.7. For the purposes of this simulation the distribution of gas and dust will be taken as coincident. A local but significant variation in the ISM properties is the apparently very low value of extinction within 40pc of the sun (see Burstein and Heiles 1978 for a review), this feature was modelled by having no gas or dust within 40pc of the sun.

(c) The detailed distribution of gas and dust - i.e. cloud properties - is more controversial; considering regions other than molecular clouds (where extinction is measured in magnitudes) the traditional picture is of a distribution of clouds, predominantly of one size with considerable dimensions and peak extinctions (radius ~ 5pc, extinction Eb-v=0.07) - the Spitzer "standard cloud" (Spitzer 1978; chapter 7). More recently evidence that smaller clouds exist has come from HI and statistical colour surveys (Knude 1979, 1981, Kalberla et al. 1980b), theoretical work (e.g. McKee and Ostriker 1977) also indicates that clouds of small dimensions may be expected. There are clearly strong selection effects favouring the detection and study of large obvious clouds. Throughout the simulations it is assumed that clouds are spherical and of uniform density. For an ISM in pressure equilibrium, and with the very small optical depths and masses of the clouds this assumption should not be grossly in error. Cloud models incorporating diffuse outer halos (after McKee and Ostriker 1977) were not significantly different from models

without halos. Details of the range of each parameter investigated are given in Table 8.1.

Table 8.1

Stellar distribution:

Diffuse component (10% total) scale-height 1500pc
Compact component (90% total) scale-height 250-350pc

Luminosity function:

Magnitude range $0 < m_V < 16$

Functional form: Reid and Gilmore (1982)
Weilen (1974)
Luyten (1968)

Dust distribution:

Coincident with gas distribution

gas to dust ratio assumed constant

Scale-height 80-150pc

Cloud radii 0.3-5pc

Cloud shape assumed spherical, density uniform

Cloud radii spectra proportional to $\text{radius}^{-2.5}$ – radius^{-4}

Cloud density 10-80 atoms per cubic centimeter

HI to extinction ratio 5.4E21

Void 40pc radius about sun with no dust or gas

8.6 Extinction Variations and the Distribution of Galaxies

A series of simulations investigated which extinction models resulted in severe modification of the intrinsic galaxy clustering. In particular the affects on the amplitude and form of the covariance function were considered. The affect of a uniform density spherical cloud on the galaxy distribution is to reduce the apparent galaxy surface density;

the projected optical depth decreases very slowly initially, before a rapid drop once the line of sight extinction reaches 70% of its peak. This results in a slowly decreasing covariance function to a scale corresponding to a substantial fraction of the cloud radius followed by a steep turndown. A clumpy ISM might be expected to enhance the observed clustering at small scales followed by a sharp break - cf. the model for the observed covariance function discussed in Chapter 2.

The amplitude of the galaxy clustering induced by variable extinction changes little between samples limited at 19 $<m_J<22$ as the slope of the number counts does not alter substantially (slope change of 0.1 is observed), but as already pointed out the amplitude of intrinsic galaxy clustering drops rapidly - factor of 10. The situation for eye selected catalogues is not as simple because of the selection procedures dependence on isophotal diameters (Phillipps et al. 1981). In considering the affect of variable extinction for the samples under discussion the worst possible case is examined - i.e. the affect on a 22nd magnitude sample. Fig. 8.1 illustrates the covariance function for four models; each with a mean absorption A_B=0.1 and constant extinction per unit volume inside a cloud, the cloud radii correspond to 0.5, 1, 2 and 5pc, equivalent to peak extinctions of 0.01, 0.02, 0.04 and 0.1. A power-law model for the galaxy covariance function is shown with the predicted amplitude for a 22nd magnitude galaxy sample - see Chapter 9. For a given mean extinction the two variables of interest are the cloud radius and density (effectively extinction per unit volume). For models with mean absorptions of A_B =0.1, 0.2 and 0.4 magnitudes, Fig. 8.2 illustrates the affect on the galaxy covariance function as a function of these two variables. The cloud density axis in Fig. 8.2 is calibrated in atoms per cubic centimetre of HI, the gas to Eb-v ratio used throughout is 5.4E21. The amplitude of the affect is arbritarily defined

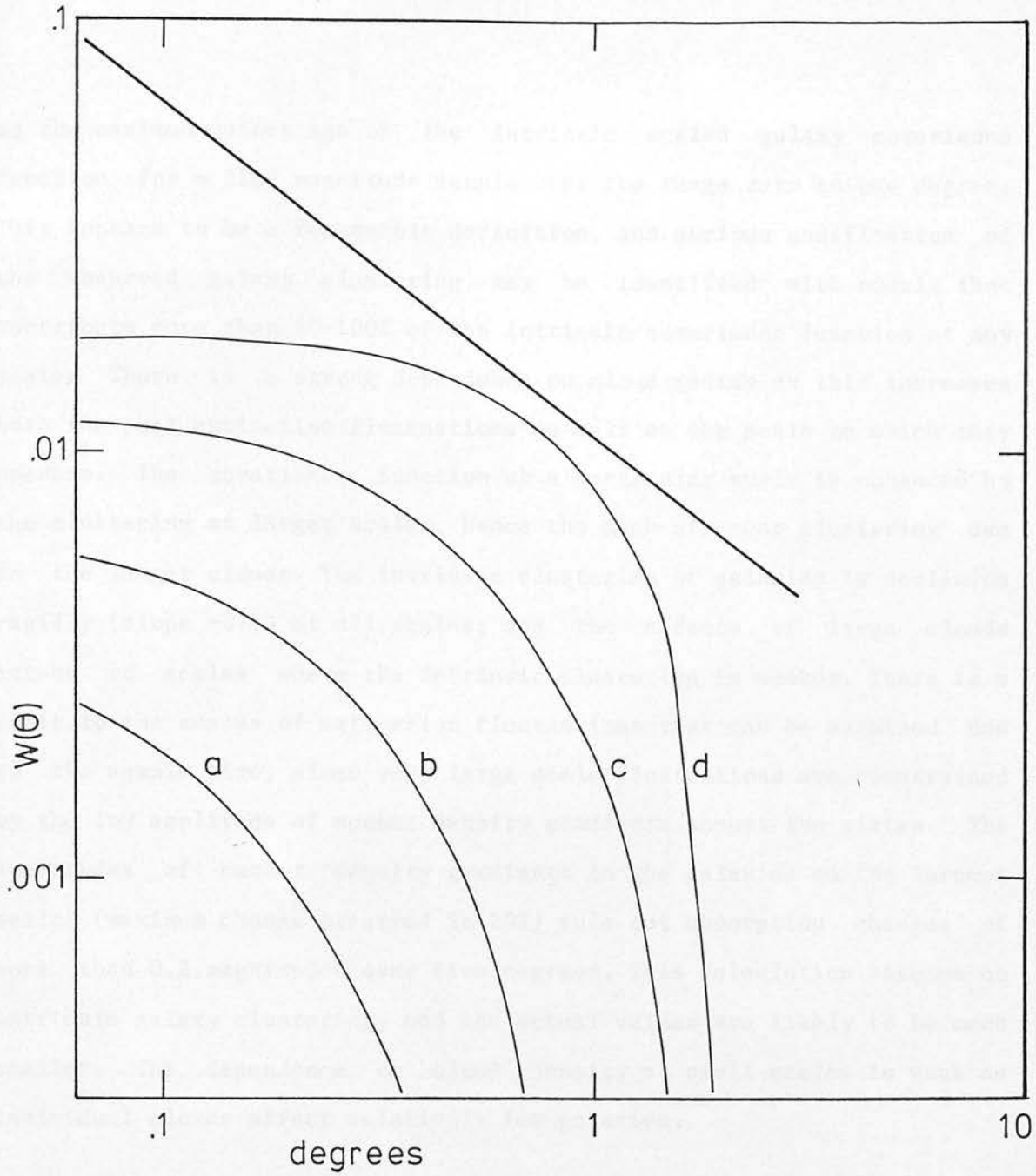


Fig. 8.1 Induced covariance function for four extinction models for a 22nd magnitude galaxy sample, together with the scaled covariance function due to intrinsic galaxy clustering. The mean value of absorbtion is A_B =0.1, the cloud density is 20 atoms per cubic centimeter and the the cloud radii are 0.5, 1.0, 2.0 and 5.0pc corresponding to curves a, b, c, and d respectively.

as the maximum percentage of the intrinsic scaled galaxy covariance function for a 22nd magnitude sample over the range zero to one degree. This appears to be a reasonable definition, and serious modification of the observed galaxy clustering may be identified with models that contribute more than 50-100% of the intrinsic covariance function at any scale. There is a strong dependence on cloud radius as this increases both the peak extinction fluctuations as well as the scale on which they operate. The covariance function at a particular scale is enhanced by the clustering at larger scales, hence the much stronger clustering due to the larger clouds. The intrinsic clustering of galaxies is declining rapidly (slope -0.8) at all scales, and the affects of large clouds extend to scales where the intrinsic clustering is weaker. There is a limit to the scales of extinction fluctuations that can be examined due to the sample size, since very large scale fluctuations are constrained by the low amplitude of number density gradients across the plates. The amplitudes of number density gradients in the galaxies on the largest scales (maximum change observed is 20%) rule out absorption changes of more than 0.2.magnitudes over five degrees. This calculation assumes no intrinsic galaxy clustering, and the actual values are likely to be much smaller. The dependence on cloud density at small scales is weak as individual clouds affect relatively few galaxies.

It is clear from Fig. 8.2 that for mean values of extinction at high latitudes in accord with observations (Section 8.2), very large high density clouds are necessary for significant modification of the intrinsic galaxy clustering. A Spitzer "standard cloud" has radius 5pc, density 12 atoms per cubic centimeter in Fig. 8.2. For mean extinction values two or three times the best observational estimates (i.e. A_B=0.2) then Spitzer type clouds are required for significant modification. If an extreme model for the value of the extinction at the poles is taken

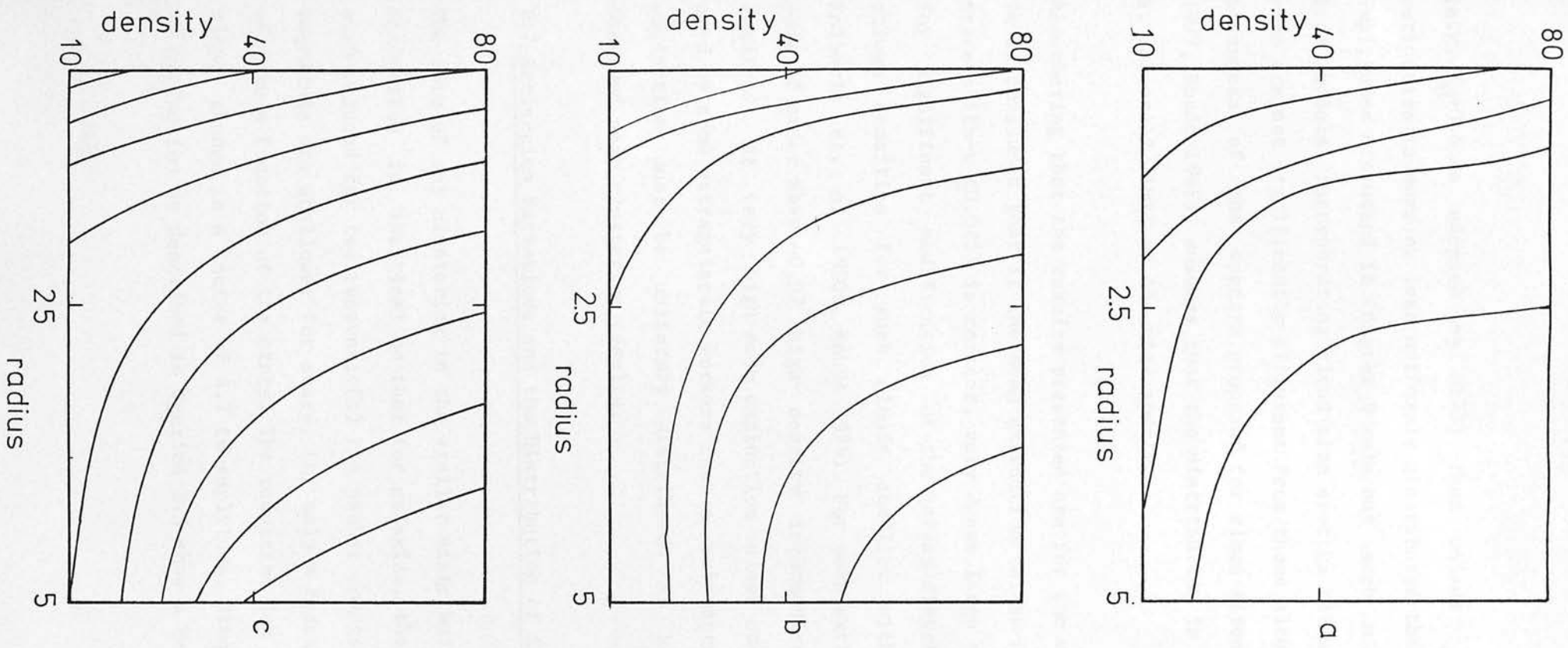


Fig. 8.2 The amplitude of the effect of extinction models on a 22nd magnitude galaxy sample. Data are for the J or B passband; panel (a) has A_B =0.1, (b) A_B =0.2 and (c) A_B =0.4. The density axis is calibrated in atoms per cubic centimeter, the radius axis is the cloud radius in parsecs. The lowest contour interval in each panel (at lower left) corresponds to 10% the second to 20%, third 40% ... The definition of the percentage changes is given in the text.

(e.g. A_B=0.6 as adopted by GP77) then unless $\geq$70% of the total extinction is more or less uniformly distributed the observed clustering amplitudes discussed in Chapter 9 rule out such a model. Conclusions from models incorporating cloud size spectra rather than a fixed cloud size are not significantly different from those already discussed - the steepness of the spectra proposed for cloud sizes (McKee and Ostriker 1977, Knude 1981), ensures that the distribution is dominated by clouds at the small limit of the size spectrum.

Remembering that the results presented are for the worst possible case, it is concluded that if the mean extinction values derived from galactic tracers (Eb-v $\leq$0.025) is correct, very dense large clouds are required for significant modification of the galaxy distribution - the high HI column densities for such clouds conflict with observation (e.g. Kalberla et. al 1980b, Knude 1979). For mean extinction values at the pole of order Eb-v =0.07 high density intermediate size clouds are required. At very high mean extinction values corresponding to A_B>0.5 used by some extragalactic workers then a substantial portion of the extinction must be uniformly distributed to be consistent with the observed deep clustering results.

8.7 Extinction Variations and the Distribution of Stars

The form of any clustering in the stellar distribution due to variable extinction is identical to that for galaxies, the amplitude however is much reduced for two reasons; (a) the number counts as a function of magnitude are shallower for stars, (b) only a fraction of the extinction affects a fraction of the stars. The reduction due to the number count slope change is a factor of 6.7 in amplitude. Simulation of the affects of extinction as described in Section 8.5 show a reduction of a factor

3.2 occurs because of the latter affect (figures for a 22nd magnitude galaxy sample, 20th magnitude stellar sample (scale-heights 300/1500pc), ISM scale-height 100pc). The amplitude of the induced clustering in stellar samples at 0.1 degrees is listed for a number of extinction models in Table 8.2 - the values are extremely small.

Table 8.2

Mean absorption A_B =0.1					
cloud radius (pc)		0.5	1.0	2.0	5.0
density (atoms cm)	10	<1E-4	1E-4	3E-4	5E-4
	20	1E-4	3E-4	5E-4	9E-4
	40	4E-4	6E-4	1E-3	1.2E-3
	80	5E-4	9E-4	1.2E-3	1.6E-3
Mean absorption A_B =0.2					
cloud radius (pc)		0.5	1.0	2.0	5.0
density (atoms cm)	10	2E-4	4E-4	6E-4	9E-4
	20	3E-4	6E-4	9E-4	1.2E-3
	40	5E-4	1E-3	1.5E-3	3E-3
	80	9E-4	2E-3	3E-3	5E-3
Mean absorption A_B =0.4					
cloud radius (pc)		0.5	1.0	2.0	5.0
density (atoms cm)	10	3E-4	5E-4	1.1E-3	2.3E-4
	20	5E-4	1.1E-3	2.3E-3	2.8E-3
	40	9E-4	1.9E-3	2.6E-3	4E-3
	80	2.2E-3	2.9E-3	4.2E-3	7E-3

(a) errors are $\pm$1E-4 throughout

The covariance functions for the deepest stellar samples for the five fields described in Chapter 7 are shown in Fig. 8.3, the surface densities and corresponding magnitude limits are given in Table 8.3.

Table 8.3

Field	Stellar density	Magnitude limit	Density gradient	Covariance amplitude
SGP	720	18.5	<2%	<3E-4
MTF	1580	18.5	20%	<3E-4
297	920	20.0	8%	<1E-3
350	1010	20.1	20%	<1E-3
352	960	20.0	9%	<3E-4

(a) number densities are per square degree, errors better than poisson. Magnitude limits to $\pm$0.2, density gradients accurate to $\pm$2%. Covariance amplitude errors $\pm$20%.

Table 8.4

Field	Observed cross-correlation	Predicted cross-correlation	Covariance upper limits
SGP	<5E-4	1.4E-3	1E-4
MTF	<1E-3	1.4E-3	3E-4
297	<5E-4	4.6E-3	1E-4
350	<1E-3	4.6E-3	3E-4
352	<5E-4	2.3E-3	1E-4

(a) observed cross-correlations and predicted cross-correlation errors $\pm$20%. Upper limits to stellar covariance function $\pm$30%.

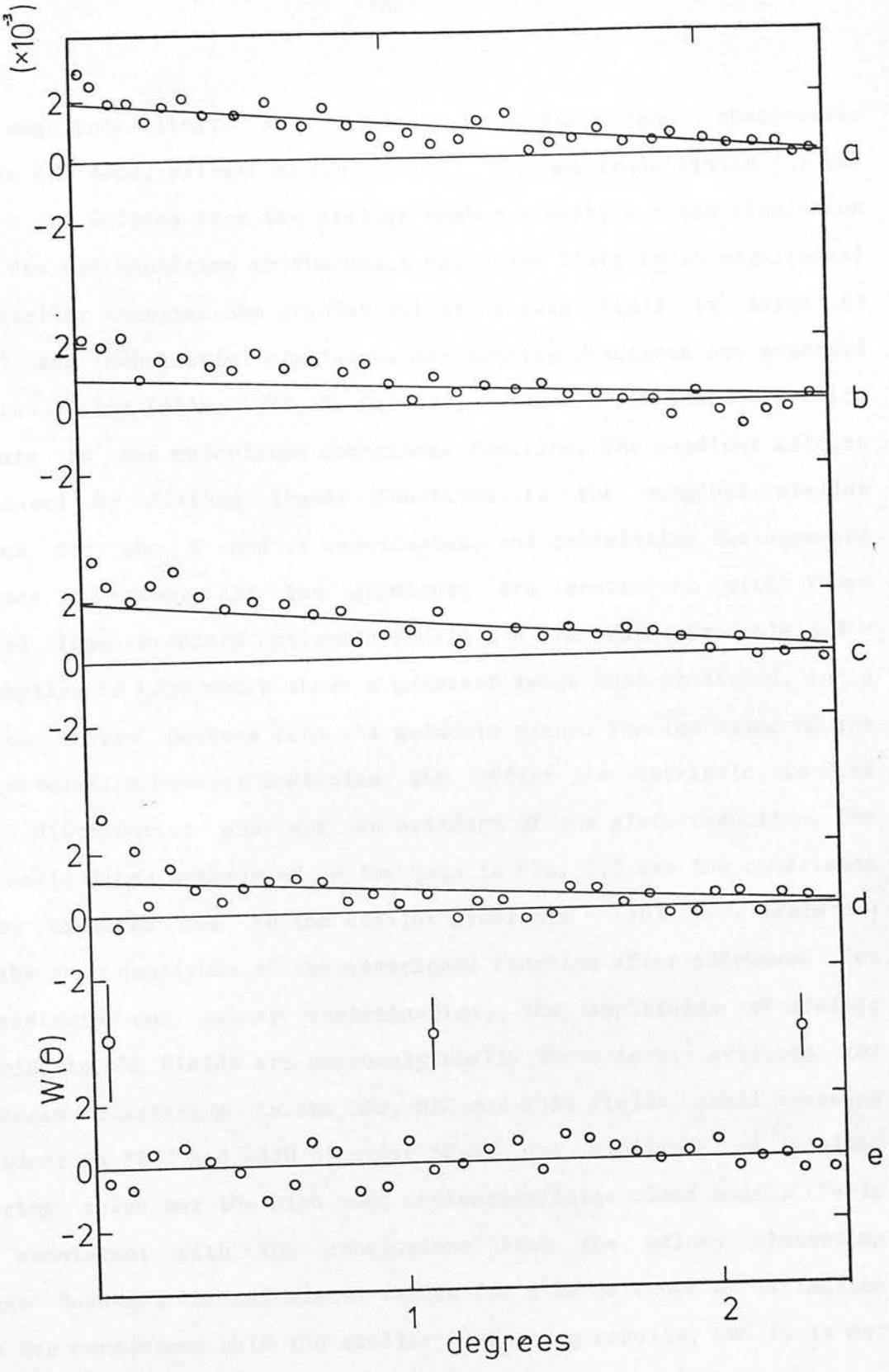


Fig. 8.3 the observed covariance functions for the stellar samples listed in Table 8.3. the solid lines are the covariance functions due to the large scale galactic gradients in the stellar samples - Table 8.3. Panel (a) MTF, (b) F297, (c) F350, (d) F352 and (e) SGP.

The V magnitude limits are reliably determined from photometric standards (N. Reid, private communication), the magnitude limits for the J samples are derived from the stellar number densities - the simulation results are not sensitive to the exact magnitude limit ($\pm$0.5 magnitudes) of the stellar samples. The angular extent of each field is large (5 degrees) and substantial stellar number density gradients are expected across the fields (Allen 1973, G. Gilmore, private communication), which contribute to the calculated covariance function. The gradient affects were removed by fitting linear functions to the marginal stellar densities for the X and Y coordinates, and calculating the expected covariance function. All the gradients are consistent with those predicted from standard galactic models and are listed in Table 8.3 - the exception is F350 which shows a gradient twice that predicted, in a direction thirty degrees from the galactic plane. The low value of the cross-correlation however indicates the affect is intrinsic to the stellar distribution and not an artifact of the plate reduction. The smooth solid lines superposed on the data in Fig. 8.3 are the covariance function expected due to the stellar gradients - Table 8.3. Table 8.4 lists the peak amplitude of the covariance function after allowance for the gradients and galaxy contamination, the amplitudes of stellar clustering in the fields are extremely small. There is no evidence for any excess clustering in the SGP, MTF and F352 fields, small excesses are evident in F297 and F350 of order 5E-4. The amplitude of stellar clustering rules out the high mean extinction/large cloud models (Table 8.2), consistent with the conclusions from the galaxy clustering results. However, the calculated values for a large range of extinction models are consistent with the stellar clustering results, and it is not clear whether the observed clustering is due to variable extinction or intrinsic to the stellar distribution.

Further constraints can be placed on the amplitude of extinction fluctuations by considering the cross-correlation between the stellar and galaxy distributions; if the stellar clustering is induced by variable extinction then the galaxies will exhibit the same clustering. The amplitude of the clustering in the galaxy samples will be a factor 22 greater than the stellar covariance function, and the amplitude of the cross-correlation between the stellar and galaxy samples should then be equal to the square-root of the product of the stellar and galaxy covariance functions - factor 4.6 larger than the stellar covariance function. If there is no extinction component in the stellar clustering then the cross-correlation should be zero. In practice because of the large fraction of the "galaxy" samples composed of stars the cross-correlation will be equal to the covariance function of the stellar sample scaled by the fraction of the "galaxy" sample composed of stars, and the fraction of the "stellar" sample composed of galaxies. Fig. 8.4 illustrates the cross-correlation functions for the stellar and deep galaxy samples in each field, Table 8.4 lists (a) the peak systematic excess of the observed cross-correlation over the calculated cross-correlation, (b) the amplitude of the cross-correlation predicted assuming the observed excess clustering in the stellar distribution is entirely due to variable extinction, and (c) upper limits on the amplitude of the stellar covariance function due to variable extinction calculated from (a). Comparison of column (c) in Table 8.4 and Table 8.2 shows that a large range of extinction models are inconsistent with the data. Limits for the V data are not as strict as the J data - values in Table 8.2 should be multiplied by 0.75 to give predicted ampitudes for V data. The amplitudes of the observed cross-correlations are all very close to those assuming the stellar clustering is intrinsic. None of the allowed models severely affect the galaxy clustering in the deep samples of Chapter 9. The data also suggest that smaller clouds predominate at

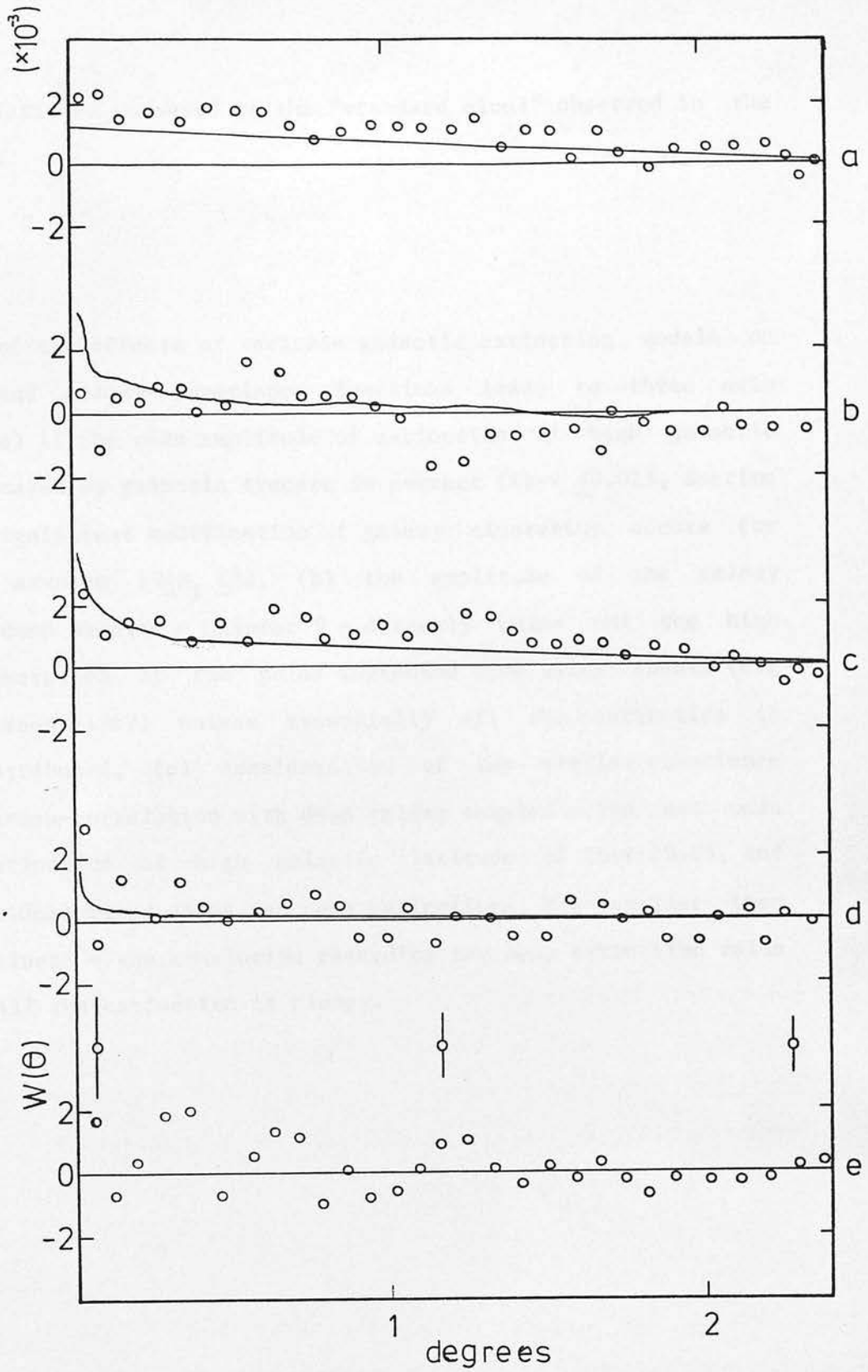


Fig. 8.4 The observed cross-correlation for the stellar samples of Fig. 8.3 and the companion galaxy samples. The solid lines are the expected cross-correlations due to the contamination of the "stellar" and "galaxy" samples by galaxies and stars respectively. Panel (a) MTF, (b) F297, (c) F350, (d) F352 and (e) SGP.

high galactic latitude compared to the "standard cloud" observed in the galactic plane.

8.8 Summary

Consideration of the affects of variable galactic extinction models on the stellar and galaxy covariance functions leads to three main conclusions; (a) if the mean amplitude of extinction at high galactic latitudes indicated by galactic tracers is correct ($E_{b-v} \leq 0.025$, Section 8.2) then no significant modification of galaxy clustering occurs for narrow angle samples $19 \leq m_J \leq 22$, (b) the amplitude of the galaxy clustering in deep samples - Chapter 9 - directly rules out the high values of absorption at the poles suggested from galaxy counts (e.g Shane and Wirtanen 1967) unless essentially all the extinction is uniformly distributed, (c) consideration of the stellar covariance function and cross-correlation with deep galaxy samples rules out mean values of extinction at high galactic latitude of $E_{b-v} > 0.05$, and suggests individual cloud sizes and peak extinctions are smaller than "standard" values - the conclusion regarding the mean extinction value assumes that all the extinction is clumpy.

9 Observations and Results

9.1 Sample Limitations

Three factors conspire to limit the information on galaxy clustering from single Schmidt plates: (a) limited number of objects, (b) small effective linear extents and (c) small magnitude range to test scaling predictions. Two fields - the V plate data - each contain approximately 12,000 galaxies, corresponding to a limiting magnitude of $m_J \sim 20.5$, the three J fields each contain 50,000 galaxies, corresponding to a limiting magnitude of $m_J \sim 22.0$. Strong clustering of galaxies on small ($\leq 5h^{-1}$ Mpc) scales is well established, and the number of independent points in the samples is much reduced. The linear scale of clustering that may be examined depends on the selection function for the sample: for the data under discussion the linear extents of the deepest samples at the modal redshift (of the n(galaxy) vs. redshift relation) are $35h^{-1}$ Mpc and $50h^{-1}$Mpc for the V and J plates respectively. The analysis of nearby galaxy and cluster samples (e.g. Seldner and Peebles 1977, GP77, Bahcall and Soneira 1983) indicate that structures - superclusters - exist on scales of at least $30h^{-1}$Mpc, a conclusion supported by more detailed redshift surveys (e.g Gregory and Thompson 1978): the samples described here are not "fair" for any but the smallest scales of galaxy clustering. The situation is compounded due to the breadth of the n(galaxy) vs. redshift selection function: the most conservative model for the deep J samples contains galaxies to redshifts exceeding $z=0.7$ and the reduction in the amplitude of the observed clustering is large - a factor of 80 between the 15th magnitude Zwicky sample and a 22nd magnitude J sample. The information making the dominant contribution to the clustering observed in each sample comes from a much narrower range in redshift space near the peak of the n(galaxy) vs. redshift relation

(e.g. Peebles 1980; section 52): the distance surveyed in redshift space is small, and it is not possible to compensate for the narrow angular sample extent by arbitrarily increasing the depth of the survey. The small angular extent of the samples results in the number of objects at bright magnitudes ($m_J<19.5$) being small - each sample contains only 1500 galaxies brighter than $m_J=19$, the limit of the Lick survey. To obtain even reasonable signal to noise in the clustering statistics from a single field the data are confined to $19.5 \leq m_J \leq 22.0$. As a result the scaling properties of the samples are expected to be poor compared to that between the Zwicky and Lick samples. As the luminosity function of galaxies is wide, extending over at least four magnitudes (e.g. Felton 1977, Davis and Huchra 1982); large clusters of galaxies in the fields result in features at fixed apparent scale - independent of magnitude. On scales exceeding small groups the scaling properties of the samples are subject to severe modification by the presence of clusters or superclusters - the deeper samples merely probing the cluster luminosity functions to fainter levels.

In spite of the limitations, the data provide significant new information on galaxy clustering, for the only published, detailed treatment of automated machine measures of deep galaxy samples is the SFEM study. SFEM investigated three fields of approximately twelve square degrees each - half the field size of the present samples - utilising COSMOS data of a much earlier epoch; important results concerning the scaling properties of small scale clustering confirmed and extended the work of Peebles and collaborators. The present samples examine an area three times that of SFEM with sample completeness, image classification probabilities and noise frequency well determined - Chapter 7.

Recently Einasto et al. (1980) and others have claimed the observed distribution of galaxies is well modelled by the presence of virtually all galaxies in a filamentary or cellular structure, the filaments occupying an extremely small fraction of space - estimates of from one to four percent are typical (e.g. Chincarini et al. 1981), corresponding to galaxy overdensities in the filaments of up to one-hundred. In the heirarchical picture suggested by the form of the low order correlation functions (Peebles 1980), taking the standard model for the spatial correlation function (equation 2.2) the average overdensity within a radius of $10h^{-1}$ Mpc of a galaxy is 0.5, and 0.14 within $20h^{-1}$Mpc. These results assume the power-law model holds to at least $20h^{-1}$ Mpc, and any break from the model at smaller scales, as proposed by SFEM, for example, decreases these overdensities. The contrast between the two models is remarkable, and the fact that both are apparently tenable is due to the strong shape dependence in the Einasto model, and the difficulty of obtaining spatial information from 2D projections. A programme to determine the limitations that could be placed on the competing models using the new data was undertaken employing monte-carlo methods to construct "synthetic" 2D samples from known spatial clustering models - this technique has been utilised by Soneira and Peebles (1978; henceforth SP78) with considerable success.

9.2 Selection Functions and Scaling

Calculations of the scaling behaviour for spatial clustering schemes using selection functions appropriate to the UKSTU/COSMOS data were made. The treatment of selection functions for deep machine measured samples has been considered in a series of pioneering papers by the Durham group (Ellis et al. 1977; henceforth EFP77, Phillipps et al. 1978, SFEM).

In contrast to the Durham analysis the primary aim of this project was not to test the detailed scaling behaviour of the covariance function and examine the galaxy number counts, but to study deep samples for evidence of large scale clustering. Investigations relying on precise magnitude information were not possible due to the lack of faint galaxy photometry in the fields; the SGP has deep galaxy photometry (Hawkins 1981), and the MTF has faint stellar photometry (N. Reid, private communication) but the three J fields possess no suitable calibration. It was deceided to consider the selection functions in terms of image number density rather than apparent magnitude. This has two main consequences; (a) all information on possible very large scale inhomogeneities is lost, and (b) the selection function is tightly constrained once a particular set of intrinsic galaxy parameters (e.g. luminosity function, K-corrections) are adopted, and a selection by number density is performed. It is essential that realistic estimates of seeing and threshold level are incorporated in the selection functions. However, samples defined by considering the brightest "n" objects were found to be relatively insensitive to reasonable changes in seeing and threshold level. This is in large part due to the extent of the COSMOS spot which effectively places a lower limit to the seeing in excess of two arcseconds. In simulations, variations of 0.75 arcseconds in seeing and 0.5 magnitudes in sky-brightness produced systematic changes in the selection function of less than 10% (i.e. more than 90% of galaxies were in common between simulations). This result was confirmed empirically by comparing object samples from J plates with very different seeings. Table 4.2 gives an excellent example, plates J3406 and J4445 have seeings that differ by at least 0.5 arcseconds yet the astronomical image samples are at least 95% complete. The small systematic effects that do operate are confined primarily to high redshift objects - which

are generally of low surface brightness - and the form of the observed clustering is not altered significantly. The computer subroutine to perform the threshold magnitude calculations for galaxies was written in collaboration with S. R. Heathcote.

There is a lack of data regarding the details of intrinsic galaxy properties (e.g. luminosity function, K-corrections, luminosity evolution). Fig 9.1 shows the number(galaxies) vs. redshift relation for in a 21st magnitude galaxy sample with a single change to the selection function: the slope parameter of the Schechter luminosity function is set at -1.25 and -0.9, spanning the range of currently favoured values (Ellis R. 1982). Galaxy luminosity evolution (which is poorly determined) modifies the selection function at high redshift enhancing the number of high redshift objects, reducing the amplitude of the observed clustering at a given limiting magnitude. All information on luminosity evolution is lost in these samples because of the lack of magnitude information. The detailed form of the selection functions for the samples discussed here are sensitive to a number of intrinsic galaxy properties which are not precisely known, and no further discussion of the details is included here. Table 9.1 lists the range of the primary selection function parameters modelled, and the components of a "standard" model for reference. Throughout, reference to a limiting magnitude for a galaxy sample refers to the total apparent magnitude at which the probability of inclusion in the sample is one half (cf. EFP77; Fig. 11).

The relation between 2D and 3D clustering has been considered in detail by Peebles (1973, 1980), Phillipps et al. (1978) and others. The inversion of 2D observations to obtain the spatial clustering is extremely unstable (Fall and Tremaine 1977), and the standard trial and

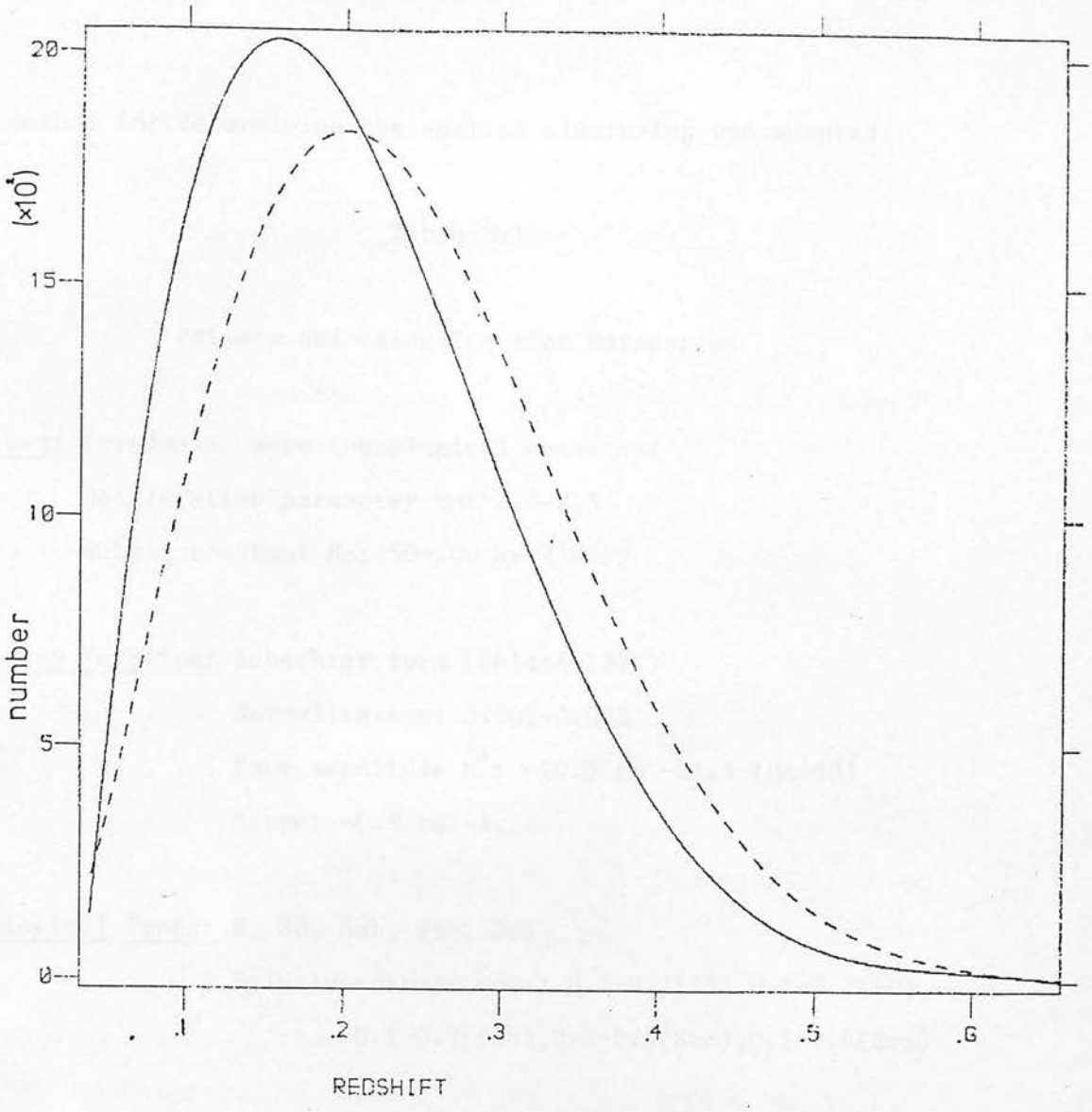


Fig. 9.1 Change in the form of the number(galaxy) vs. redshift relation due to a change in the slope of the Schechter luminosity function. The slope for the solid curve is -1.25, and for the dashed curve -0.9. The curves are normalised to give the same total number of objects.

error method for determining the spatial clustering was adopted.

Table 9.1

Primary Selection Function Parameters

Cosmology: Friedmann, zero cosmological cosnstant
Deceleration parameter Qo: 0.0-0.5
Hubble constant Ho: 50-100 km $s^{-1}Mpc^{-1}$

Luminosity Function: Schechter form (Felton 1977)
Normalisation: 0.002-0.003
Knee magnitude M^*: -20.5 to -21.5 (Ho=50)
Slope: -0.9 to -1.25

Morphological Types: E, S0, Sab, Sbc, Scd
Relative proportions: 0.1-0.2(E), 0.1-0.2(S0),
0.1-0.3(Sab),0.1-0.4(Sbc),0.1-0.4(Scd)

Profiles: exponential disc systems, and $R^{1/4}$ ellipticals
galaxies assumed pure disc or elliptical and observed face on
Profile scales taken from Pritchet and Kline (1981)

K-corrections: taken from Ellis R. (1982)
Pence (1976)
Coleman et al. (1980)

Luminosity Evolution: taken from Bruzual (1981)

Table 9.1 (cont)

Measuring Process Parameters

Seeing: Point spread function assumed gaussian
0.75-2.5 arcseconds (radius containing 90% of total light)
(COSMOS 32 micron spot contributes "seeing" of >2.0 arcseconds)

Sky-brightness: 22.5-23.0 magnitudes per square arcsecond in J
21.5-22.0 magnitudes per square arcsecond in V

COSMOS threshold level: 7-12% of sky-brightness

Magnitude errors: Model for combined COSMOS/plate errors in thresholded magnitude taken from measures of multiple plates.

Standard Model Parameters

deceleration parameter: 0.05
Hubble's constant: 50 km $s^{-1}Mpc^{-1}$
K-corrections: from Coleman et al. (1980)
luminosity evolution: exponential star formation, μ=0.7 from Bruzual (1981)
luminosity function: M^*=-21.0 slope =-1.25
morphological type proportions: .15(E) .20(S0) .20(Sab) .20(Sbc) .25(scd)
seeing: 2.0 arcseconds
sky-brightness: 22.75 magnitudes per square arcsecond
threshold level: 10%

Using NAG routines to perform the numerical integrations, the computation of scaling behaviour for any spatial clustering scheme is

relatively simple. The standard results were readily reproduced (e.g. Peebles 1980; sections 52 and 56), but it is worth stressing the effect that the behaviour of spatial clustering at large scales has on 2D projections at small scales. Fig. 9.2 shows 2D projections at 21st magnitude: in each case a power-law spatial covariance function to $10h^{-1}$ Mpc with index $\sigma = 1.8$, scale-factor $r_o = 5h^{-1}$Mpc is combined with different behaviour at large scales; (a) power-law extends to $150h^{-1}$ Mpc, (b) power-law slope changes at $10h^{-1}$Mpc to $\sigma = 2.8$, (c) as (b) but slope change to $\sigma = 3.8$, (d) slope change at $10h^{-1}$Mpc to $\sigma = 2.8$ and an anticorrelation ($\xi(r) = -0.05$) in the range r= $20\text{-}40h^{-1}$Mpc with $\xi(r)=0$ thereafter. In the case of model (d) the covariance function integrates to zero for scales greater than $40h^{-1}$Mpc.

9.3 Observations

It is shown below that the limiting factor for the data under discussion is the small sample size and consequent large fluctuations induced by the presence of large scale structure in the galaxy distribution. A number of statistical techniques were employed to examine the galaxy distributions, but none of these gave information not contained in the low order correlation functions. This is partially due to the low signal to noise of the clustering - sophisticated statistics are poorly determined. The traditional binning techniques (e.g. Masson 1979) gave very little useful information, and as the current debate concerning galaxy distributions revolves around the form and effectiveness of the correlation functions, the results presented are confined to the low order correlation functions.

The covariance function for all objects (to the limit of the faintest sample - Table 7.6) in the SGP field calculated from two different

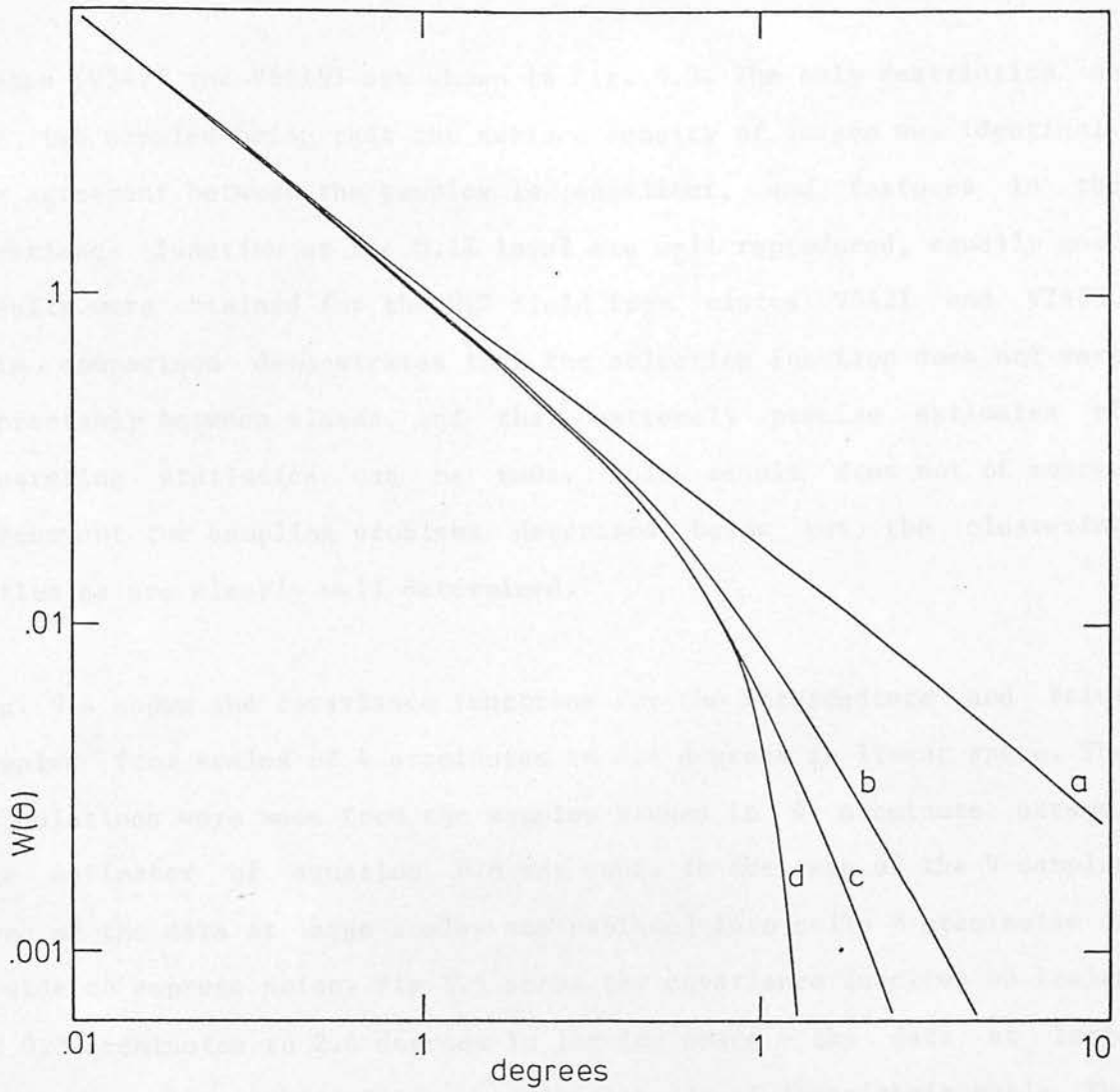


Fig. 9.2 Change in the form of the observed covariance function for a 21st magnitude galaxy sample due to different behaviour of spatial clustering at large scales. Each curve corresponds to different behaviour beyond $10h^{-1}$Mpc, with the same power-law covariance function at smaller scales. The models corresponding to each curve are described in the text.

plates (V3475 and V6619) are shown in Fig. 9.3. The only restriction on the two samples being that the surface density of images was identical. The agreement between the samples is excellent, and features in the covariance function at the 0.1% level are well reproduced, equally good results were obtained for the MTF field from plates V5421 and V2499. This comparison demonstrates that the selection function does not vary appreciably between plates, and that extremely precise estimates of clustering statistics can be made. This result does not of course circumvent the sampling problems described below but the clustering estimates are clearly well determined.

Fig. 9.4 shows the covariance functions for the intermediate and faint samples from scales of 4 arcminutes to 2.4 degrees in linear space. The calculations were made from the samples binned in 4 arcminute arrays, the estimator of equation 6.6 was used. In the case of the V samples some of the data at large scales was rebinned into cells 8 arcminutes on a side to supress noise. Fig 9.5 shows the covariance function on scales of 0.3 arcminutes to 2.4 degrees in log-log space - the data at large scales have been rebinned into equal intervals of logarithmic angle. The "galaxy" samples contain a considerable fraction of stars - Table 7.6. This results in the "galaxy" samples exhibiting clustering due to the large scale gradients in the stars due to galactic gradients - Chapter 8. The gradients are significant in the MTF and F350 fields, and the data for the MTF and F350 fields have been corrected for the stellar gradients. Otherwise no corrections have been made to the data. Estimates of the covariance function at scales less than 4 arcminutes were calculated using a monte-carlo estimator - equation (6.3). The cellular estimation procedure was extended to 1 arcminute bins to ensure no systematic differences between the estimators were present.

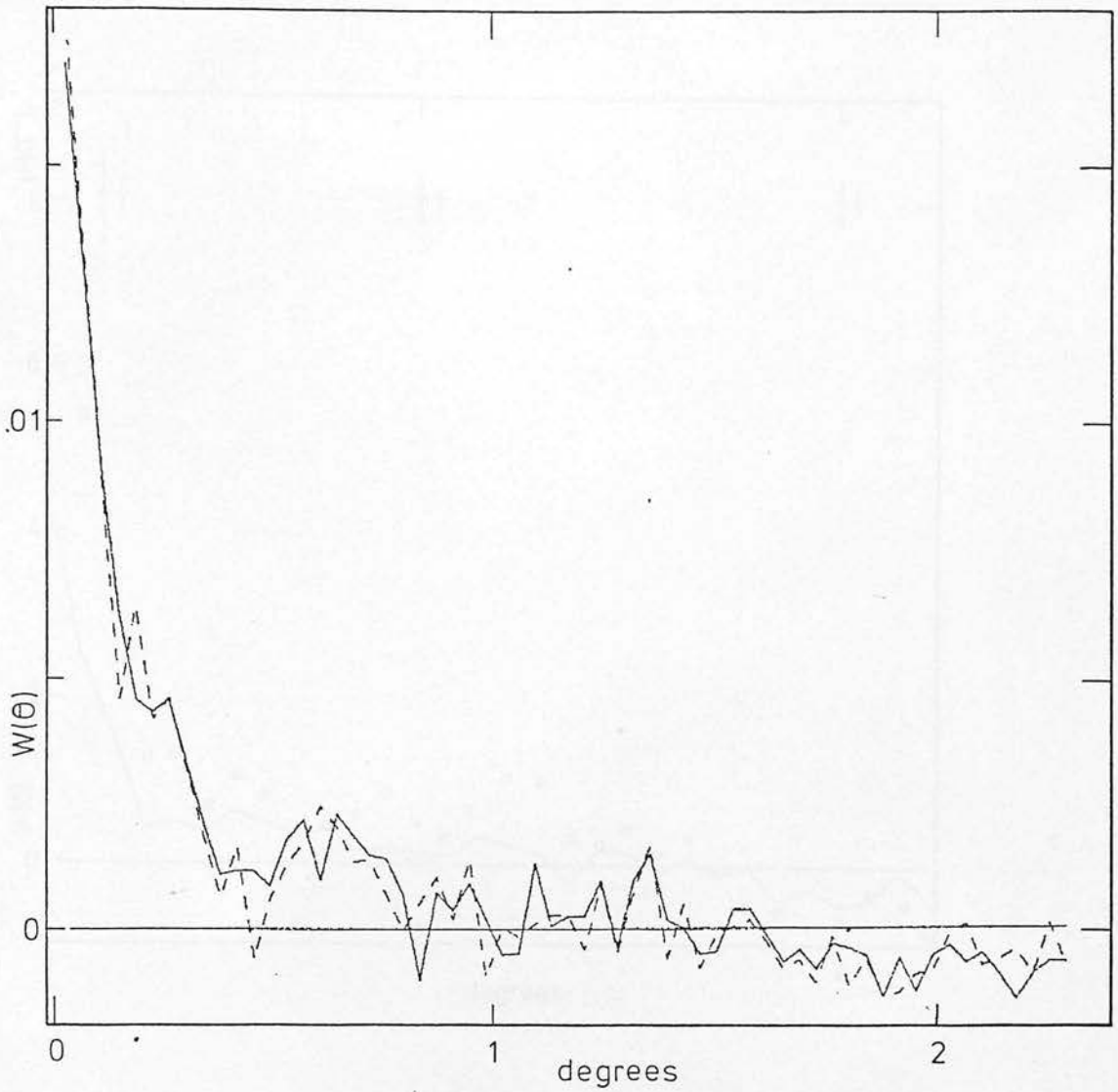


Fig. 9.3 Comparison of the covariance function for all images in the SGP field from two different plates - V3475 and V6619. The reductions were performed independently, and the only constraint on the samples being that the image surface density is identical.

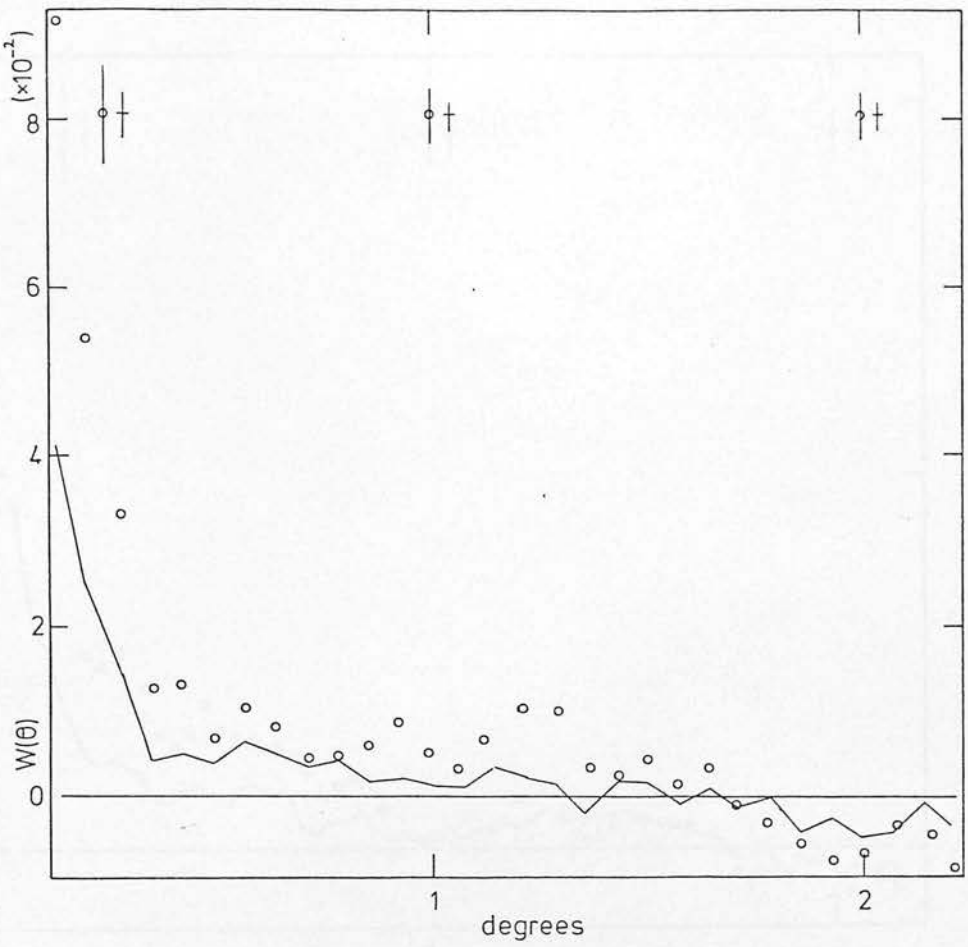


Fig. 9.4 The covariance function for all five fields at large scales, plotted in linear space. The bright samples from the J data have been omitted for clarity because of the high amplitude at small scales and noisy behaviour at large scales. In all cases the data towards the top of the figure corresponds the the lower number density sample plotted.

Fig. 9.4a the SGP sample

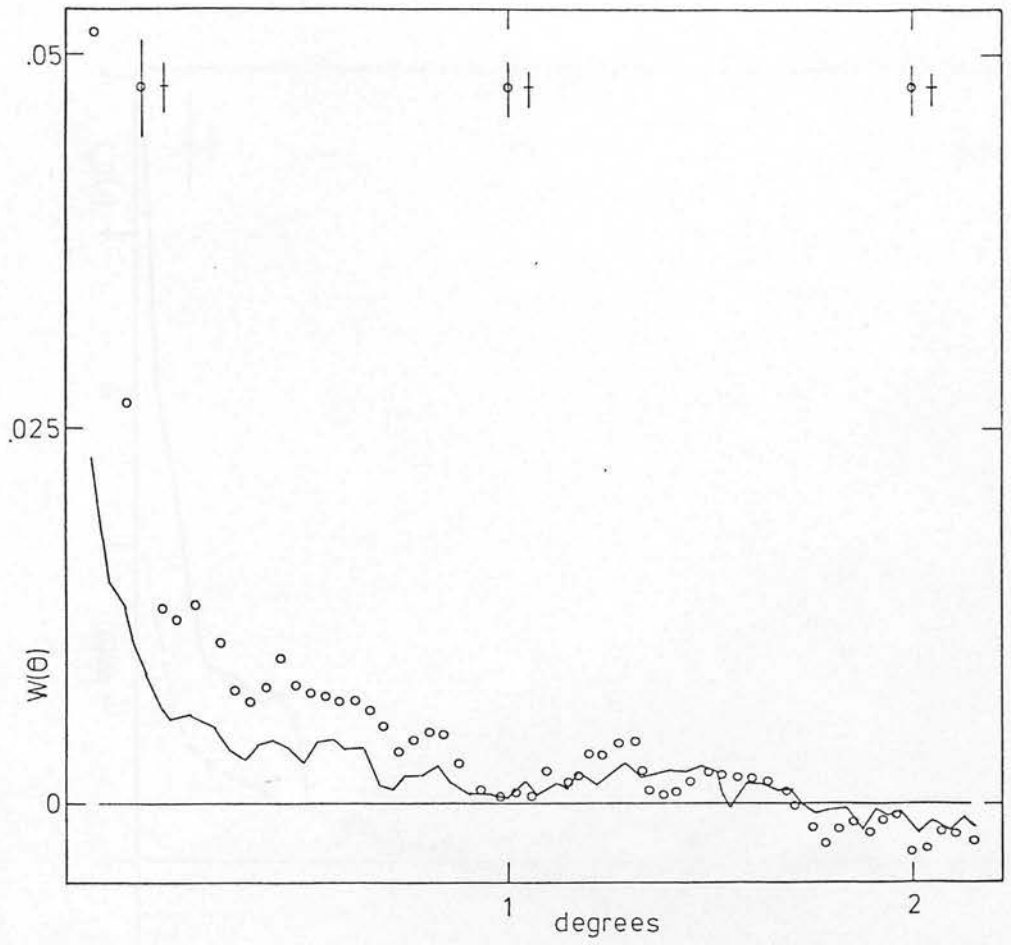


Fig. 9.4b the MTF sample

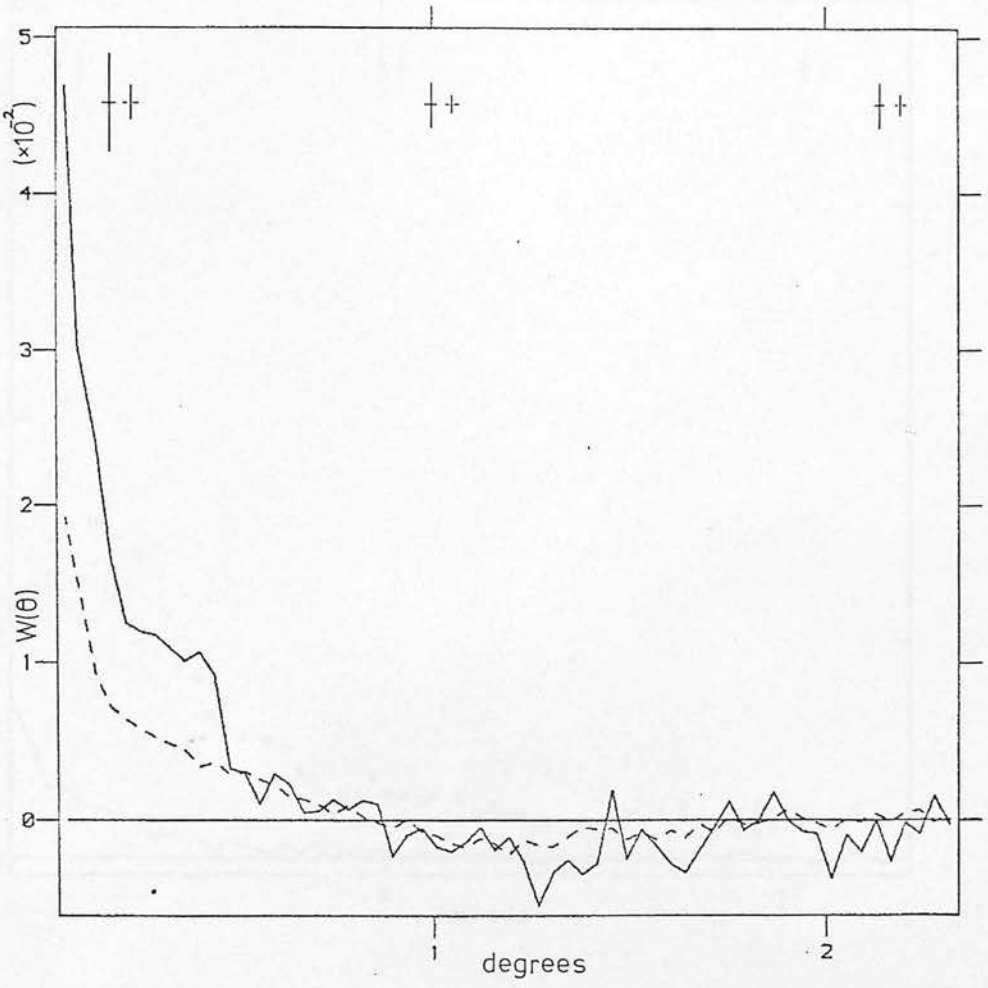


Fig. 9.4c the F297 sample

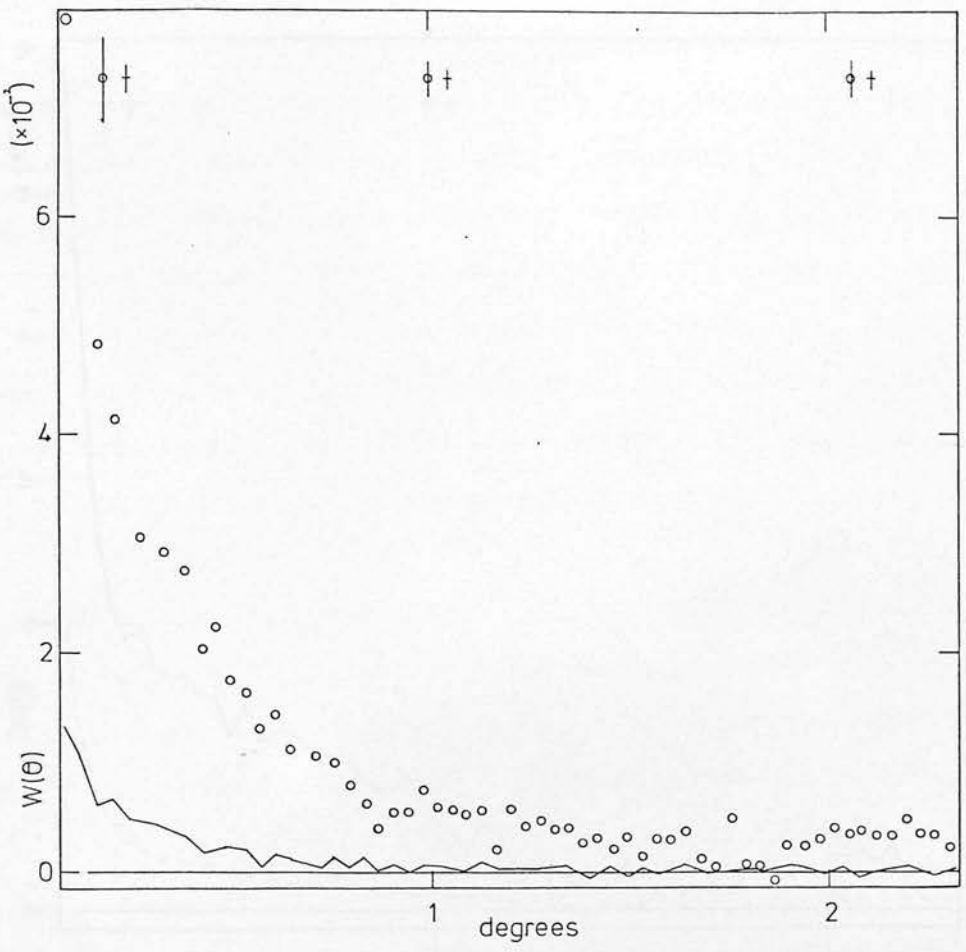


Fig. 9.4d the F350 sample

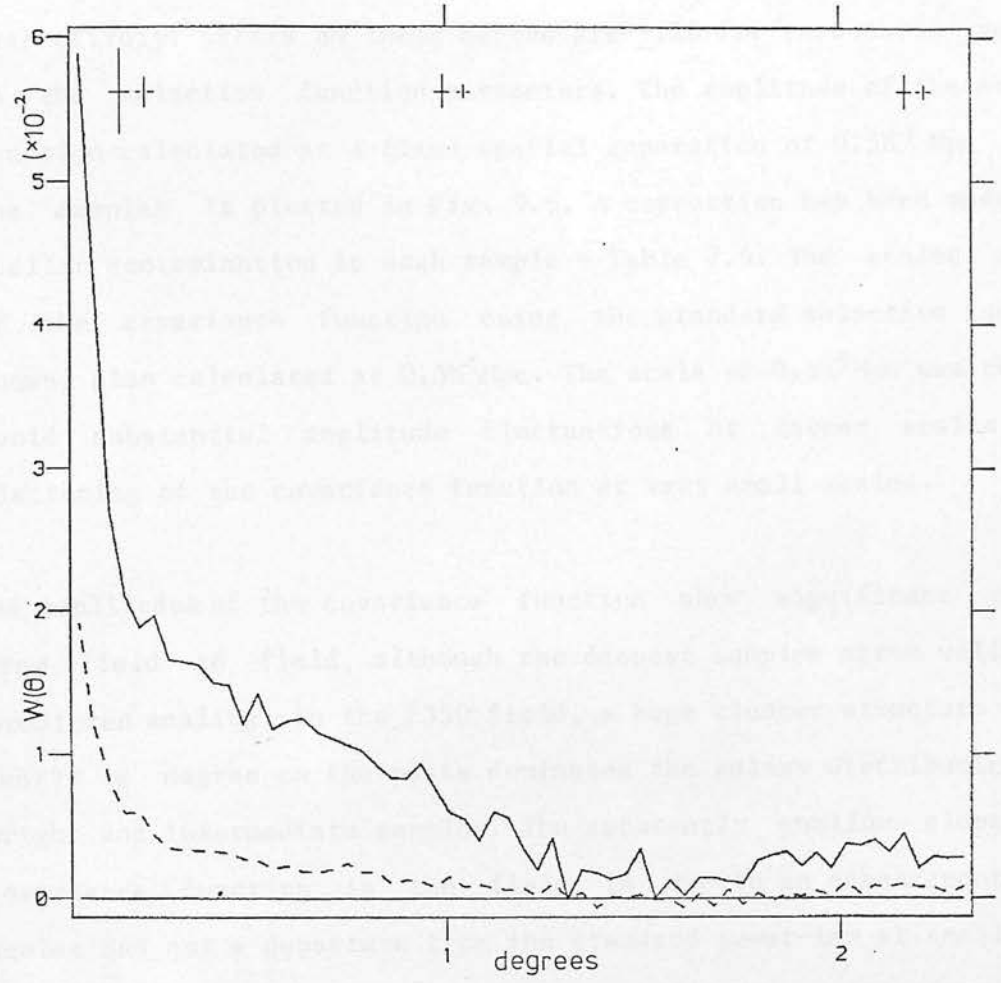


Fig. 9.4e the F352 sample

The effective depths of the J samples limited at number densities per square degree of 200, 900 and 2500 objects are 0.11, 0.19 and 0.32 respectively. Errors on these depths are $\pm$15% for reasonable variations in the selection function parameters. The amplitude of the covariance function calculated at a fixed spatial separation of $0.5h^{-1}$ Mpc for all the samples is plotted in Fig. 9.6. A correction has been made for the stellar contamination in each sample - Table 7.6. The scaled amplitude of the covariance function using the standard selection function is shown, also calculated at $0.5h^{-1}$Mpc. The scale of $0.5h^{-1}$Mpc was chosen to avoid substantial amplitude fluctuations at larger scales, and any flattening of the covariance function at very small scales.

The amplitudes of the covariance function show significant variations from field to field, although the deepest samples agree well with the predicted scaling. In the F350 field, a huge cluster structure extending nearly a degree on the plate dominates the galaxy distribution for the bright and intermediate samples. The apparently shallow slope of the covariance function in the field is due to an enhancement at large scales and not a departure from the standard power-law at small scales. The SGP V sample also shows a significant excess of well defined clusters, and a substantial portion of the departure from a single power-law form in the SGP is attributable to a clustering enhancement at intermediate scales. The error bars shown do not take account of the large field to field scatter observed in the simulations described below. It is evident that the detailed behaviour of the covariance function varies considerably from field to field - particularly at large scales. The comparison of Fig. 9.3 demonstrates these differerences are significant in the sense that they are not due to plate material, COSMOS or selection function variations.

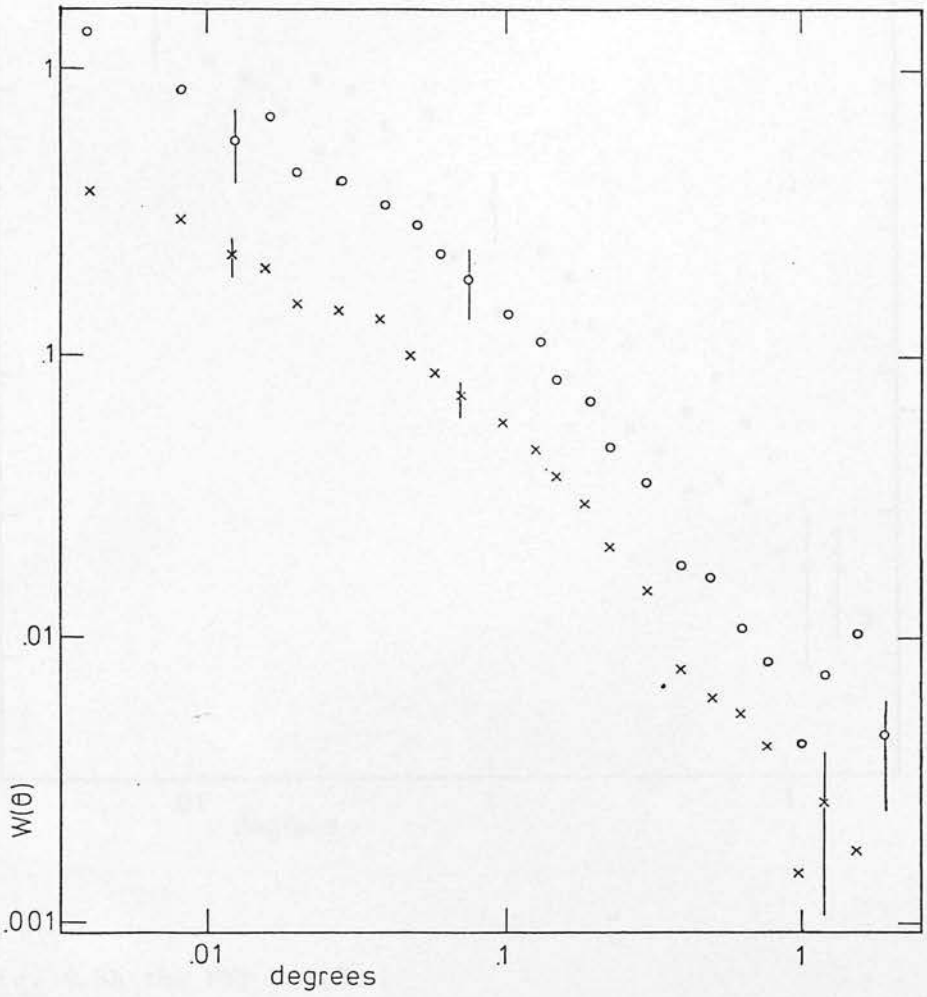


Fig. 9.5 The covariance function for all five fields at small scales, plotted in log-log space. Note the differing slopes from field to field, and the variable behaviour at very small scales. In all cases the data are for bright, intermediate and faint samples reading from top to bottom.

Fig. 9.5a the SGP sample

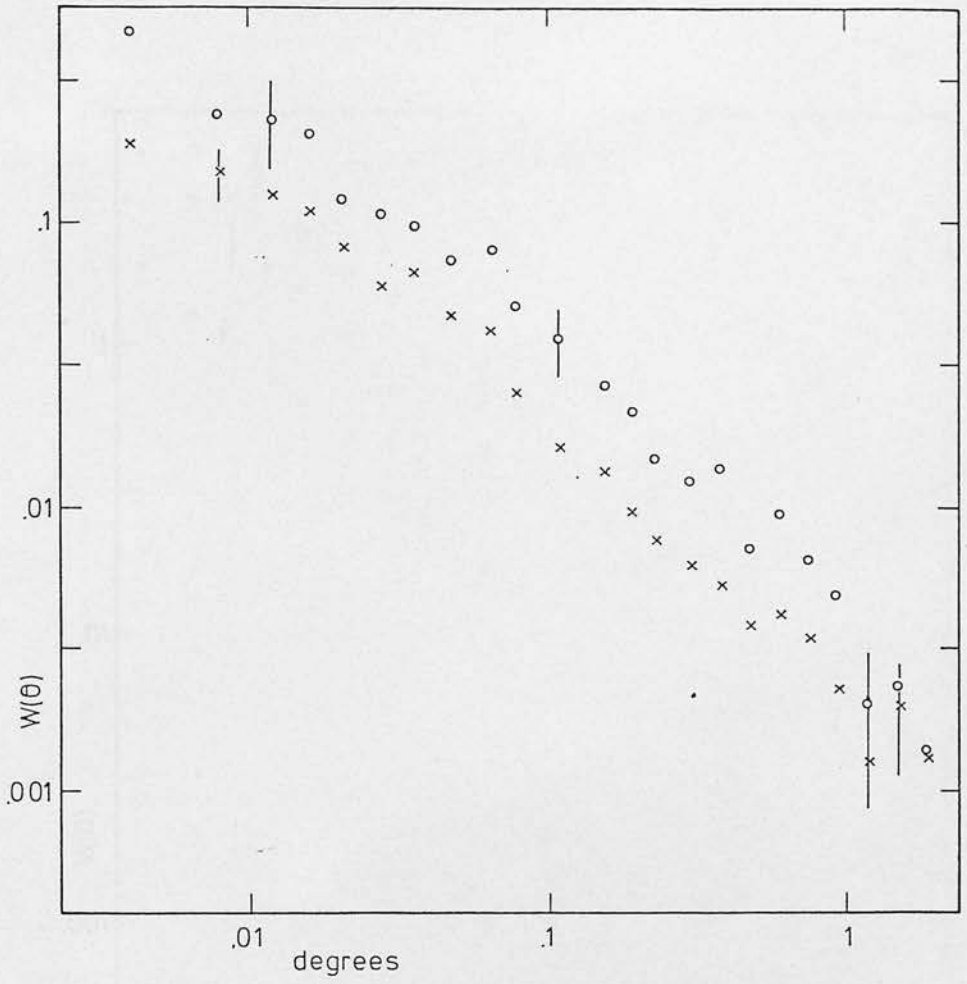


Fig. 9.5b the MTF sample

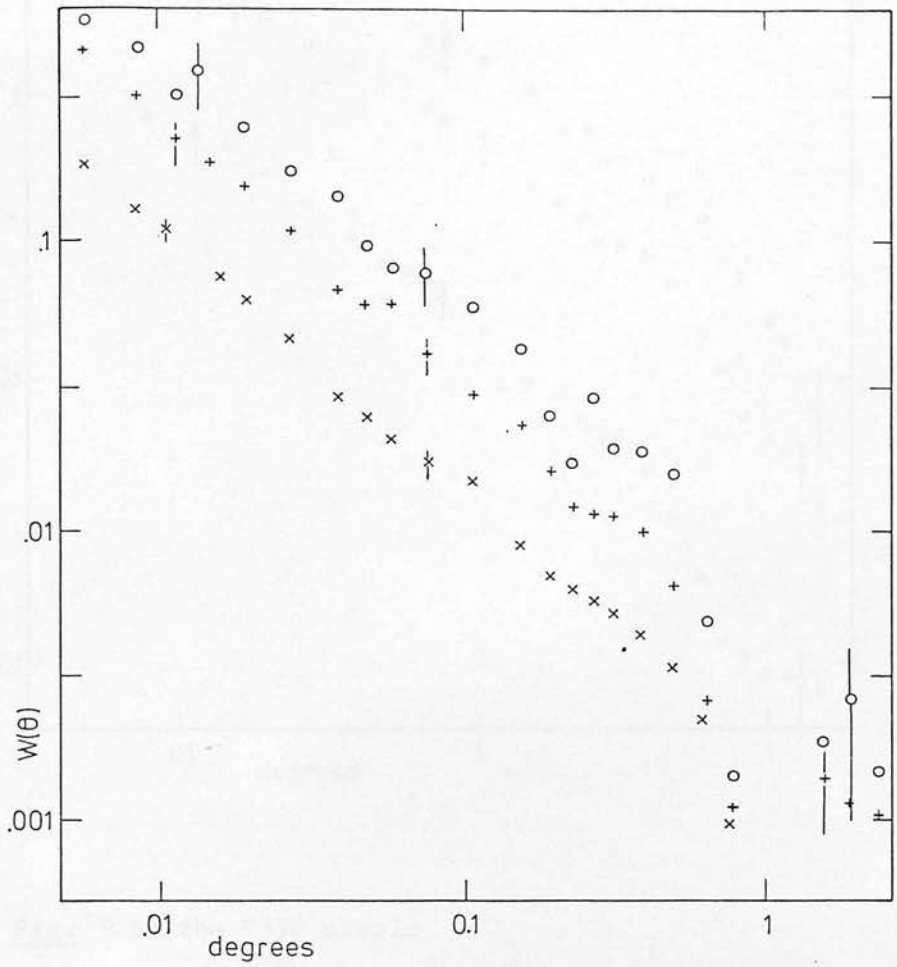


Fig. 9.5c the F297 sample

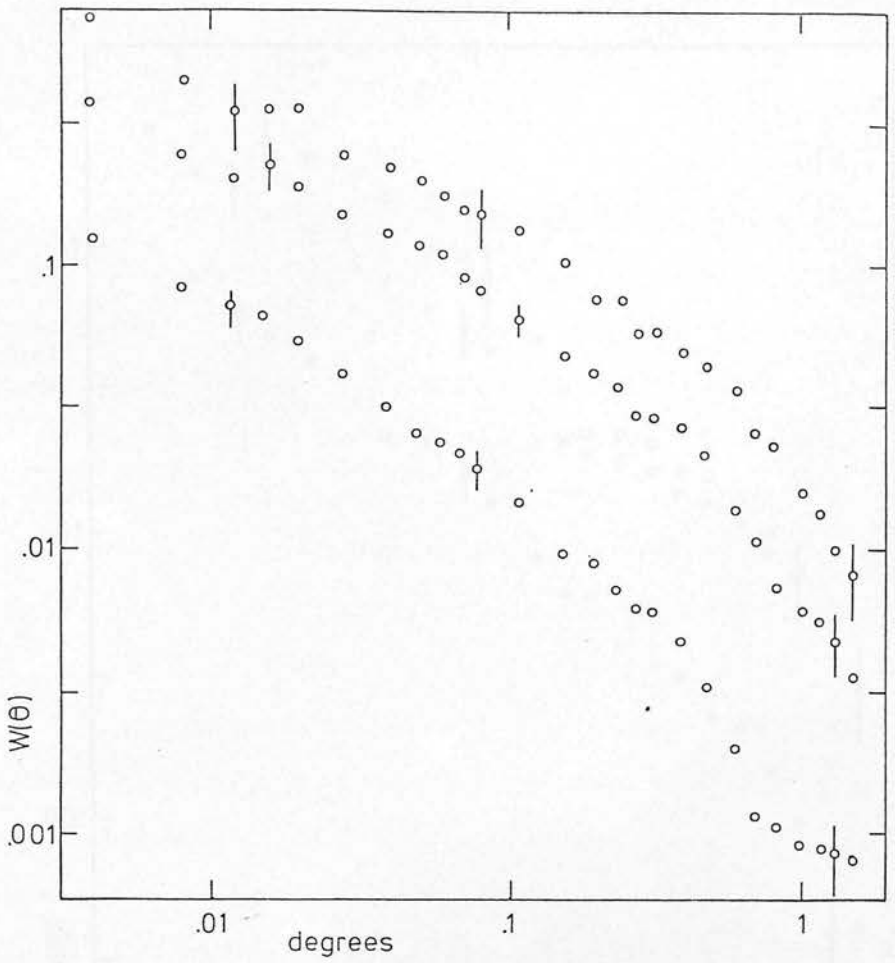


Fig. 9.5d the F350 sample

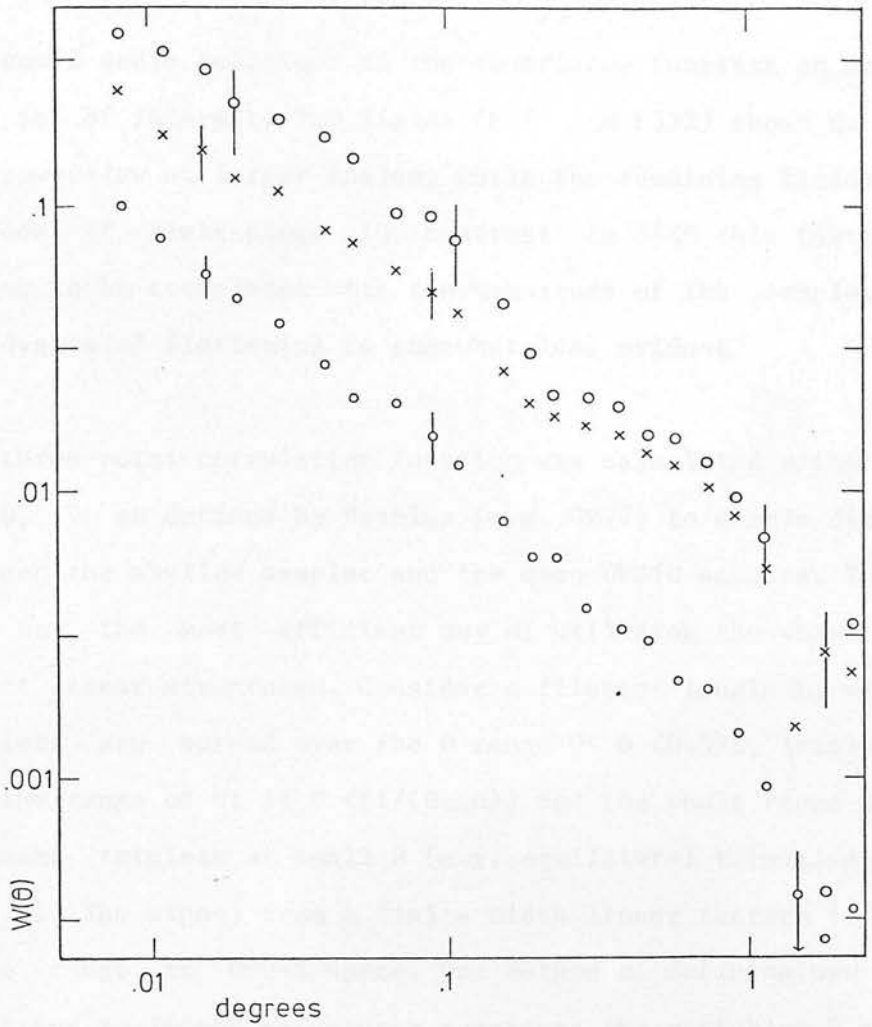


Fig. 9.5e the F352 sample

The small scale behaviour of the covariance function on scales of <0.05h Mpc is of interest. Two fields (F297 and F352) shown no departure from the power-law at larger scales, while the remaining fields show various degrees of flattening. In contrast to SFEM this flattening does not appear to be correlated with the magnitude of the sample, and overall the degree of flattening is somewhat less evident

The three-point correlation function was calculated using the variables θ, U, V as defined by Peebles (e.g. GP77) to enable direct comparison between the shallow samples and the deep UKSTU samples. These variables are not the most efficient way of utilising the shape information to detect linear structures. Consider a filament length L, width X; linear triplets are spread over the θ range $0< \theta <0.5*L$, triplets are present for the range of U; $1< U <(1/(\theta min))$ and the whole range of V contains filament triplets at small θ (e.g. equilateral triangles with side less than X). The signal from a finite width linear feature is spread over a large range in θ-U-V space. One method of defining new variables more sensitive to linear structures considers the variables D and W, where D is the maximum separation of two galaxies in the triplet and W, the perpendicular distance between the central galaxy and the line defining D. For any value of $D < L$ the filament contributes an excess of triplets up to a value Wmax (where Wmax X), as D is increased beyond L the number of triplets decreases rapidly. Redefining the variables retains the (almost unique) ability of the three point function to exactly take account of the pairwise clustering in the data.

Only small scale estimates of the three-point function are available from the data. Even at small scales the estimates of the three-point function are extremely noisy; Table 9.2 lists the values of "P" for Peebles' model of the angular three-point function in terms of the

two-point function (Peebles 1980; section 54). The values are averages over all triplets with smallest separation between galaxies of $\geq 0.4h^{-1}$Mpc and largest separation of $\leq 5h^{-1}$Mpc, the estimates have been corrected for the stellar contamination in each sample.

Table 9.2

Values of the three-point function parameter P

Field	Sample Bright	Intermediate	Faint
SGP	2.69±0.4	2.95±0.25	
MTF	2.15±0.4	1.25±0.25	
297	1.15±0.4	0.82±0.25	0.71±0.15
350	1.78±0.4	2.72±0.25	2.38±0.20
352	1.47±0.4	1.06±0.25	0.20±0.10
Jagellonian	0.78±0.2		

The field to field variation is large, within the range of separations used, the standard deviation of P is large for any sensible bin size in θ-U-V or D-W space - ±0.45 for bin size D $=1h^{-1}$Mpc and W $=0.25h^{-1}$Mpc. The estimates presented were calculated in D-W space but the results are not significantly differerent in θ-U-V space. The values of the ratio P/Q, relating the angular to spatial functions are 1.20, 1.22 and 1.24 for the number densities at which the effective depths defined above were calculated. Errors are ±10% for reasonable variations in the selection function, and the ratio shows very little dependence on sample depth. The noise precludes the detailed examination of any trends with shape for the three-point function - GP77 also found the three-point estimates were extremely noisy using a much larger data set. The mean values of Q and P are close to that predicted from the SP78, but field to field fluctuations are large. In this context it is worth noting that the

Jagellonian field is also not a "fair sample", reducing the significance of the discrepancy with the Zwicky and Lick data found by GP77. The Jagellonian field was analysed to ensure consistency between estimates, the result in Table 9.2 is in excellent agreement with that found by Peebles (GP77). Systematic variations in the value of Q in excess of 50% can be ruled out for the data for a single field, and an average over all five fields allows variations greater than 25% to be ruled out. The range of scales is limited and the constraints on possible systematic deviations from the model postulated by Peebles are weak, but the results are consistent with the observations of the shallower catalogues.

No reliable estimate of the coefficients for the model of the four point function in terms of the two-point function was possible. The technique of Bonometto and Sharp (1980) and the direct estimation of the linear combinations of the coefficients (Fry and Peebles 1978) gave wildly fluctuating results, behaviour that was also seen in the simulated fields described in the next section. The signal to noise of the four point function in the samples allows no discussion of its behaviour.

9.4 Direct Simulation of Galaxy Samples

In a more direct approach, the scaling of spatial clustering was considered by simulation of the appearance of 3D clustering schemes in two dimensions. The technique is simple; galaxies are generated in 3D space according to some input spatial clustering. The 2D sample is constructed by accepting each galaxy in the sample with a probability determined from the selection function, and testing to see whether a galaxy's projected position falls within the confines of the 2D sample. The method possesses the significant advantage that sampling variations

are taken into account directly - no homogeneity or narrow-angle assumptions are required, as in the numerical scaling calculations.

The selection functions for the simulations were identical to those for the numerical scaling calculations. The basic clustering of galaxies followed two schemes; the heirarchical model of SP78 and power-law clustering (e.g. Shanks 1979). The simulation of very large structures - filaments and sheets for instance - was accomplished by distributing galaxies according to either scheme within higher order structures. This appears to be reasonable; the small scale distribution of galaxies within filamentary structures studied so far, does not depart significantly from the standard models - the shape dependence in Einasto's and other schemes is imposed at large scales ($> 10h^{-1}$Mpc). Three large scale clustering schemes were modelled; (a) distribution of galaxies in sheet like structures, (b) distribution in linear filaments, and (c) distribution in "bubble" or "cellular" structures. In each case these units were randomly distributed, and represented the largest scale of clustering present. In the case of bubbles, a bubble center is chosen, and galaxies distributed in a shell a specified radius from the bubble center.

For the very large scale structures a key variable is the overdensity within the structures; the adiabatic theory together with numerical simulations (Doroshkevich et al. 1980) predict the existence of very large non-linear structures (in contrast to the gravitational clustering picture). The overdensity should be at least unity in the adiabatic picture; a crude estimate of the overdensity can be made by noting that the overdensity will be of order $(1+Z_{nl})$ for sheets and $(1+Z_{nl})^2$ for filaments, where Z_{nl} is the redshift at which the density becomes non-linear. Estimates of Z_{nl} of order 10 are suggested (Melott 1983,

preprint), giving values in reasonable agreement with observations (e.g. Chincarini et al. 1981). The clumpiness observed in the 2D samples is clearly dependent on the size of the structures, however, the very small projected scales of single UKSTU plates make the observations considerably less sensitive to size considerations when structures of order $30h^{-1}$Mpc or larger are present.

A number of "sophisticated" facilities are included in the simulation that allow the luminosity function and morphological type frequency to vary with enviroment for instance. The data are not sufficiently precise to warrent discussion of secondary details and results from "basic" models are described here.

SP78 develop a model around a truncated heirarchy, the largest scale on which significant structure exists is ~ $18h^{-1}$Mpc, and on a scale of $10h^{-1}$ Mpc SP78 increase the power by having only a small reduction in the heirarchy level-scale at ~ $10h^{-1}$ Mpc. This feature is necessary to reproduce the almost constant slope of the 2D covariance function from the Lick data. At very small scales the heirarchy is truncated from scales of $600h^{-1}$kpc, with no galaxies closer together than ~ $40h^{-1}$kpc. The small scale truncation appears physically reasonable (see below) although the enhancement of clustering at $10h^{-1}$ Mpc does not arise naturally from the gravitational clustering model.

Twenty simulations of the SP78 model with selection function and sample extents chosen to match the present data were run. The following conclusions were evident; (a) the slope of the 2D power-law on scales 0.1 to 0.5 degrees was 0.94 ± 0.08, with extreme variations of 0.7 and 1.2, (b) the break scale for the power-law varied by more than a factor two between samples, with a mean value of 0.6 degrees, (c) at scales

beyond 0.5 degrees variations in $\hat{W}(\theta)$ of amplitude 0.01 between samples were evident, these differences were correlated over scales of up to a degree - i.e. two UKSTU samples can show systematic differences in the amplitude of $\hat{W}(\theta)$ of 0.005-0.01 over scales of a degree, (d) negative values of $\hat{W}(\theta)$ occured in 30% of the simulations with amplitudes typically 0.005. Fig. 9.7 shows the mean covariance function for the model together with the two most extreme covariance functions from single UKSTU fields in log-log space. The Shanks (1979) model can also reproduce the overall behaviour seen in the UKSTU fields, and it is interesting to note that large scale structure ($\sim 25h^{-1}$Mpc) is also present in this model.

In the direct simulations the observed 2D power-law break scale is smaller, and power-law slope steeper than numerical scaling calculations indicate for the identical spatial clustering scheme. This is primarily due to the small size of the UKSTU fields, and the consequent undersampling of large scale fluctuations. This is an important point - the covariance function on small scales is modified by the undersampling of large scale fluctuations. The undersampling has a greater effect at bright magnitudes - where structures subtend larger angles - and reduces the apparent power-law break scale even further in these samples. This effect was also reproduced in the simulations.

In the present data, at small scales < 0.05 degrees corresponding to spatial scales of $<0.5h^{-1}$Mpc a flattening of the covariance function to a slope significantly less than the nominal value of 0.8 in two fields is observed. The Lick and Jagellonian samples provide no data on this scale due to the large cell size of the galaxy counts. Redshift information is dominated by a number of other factors at this scale, and possible selection effects in the Zwicky catalogue on small scales have already

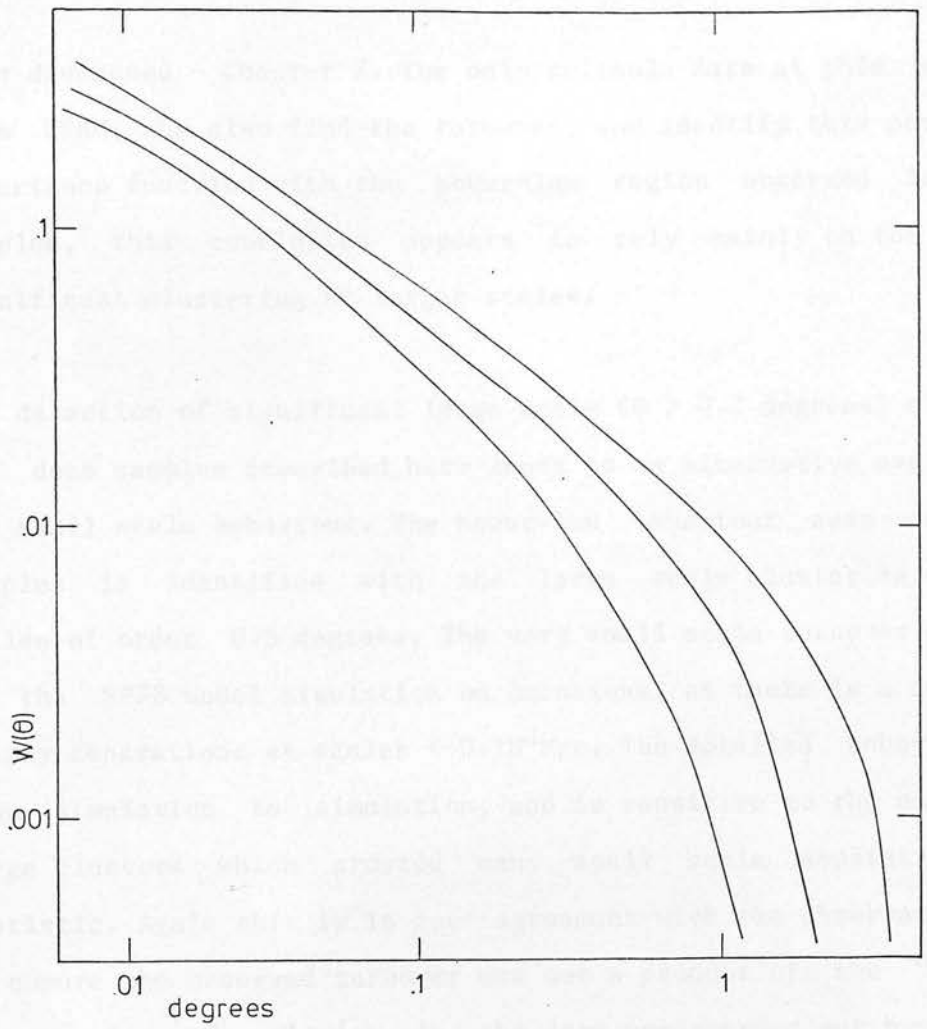


Fig. 9.7 Example covariance functions calculated from the simulations of the SP78 model described in the text. The data are for a 22nd magnitude sample. The middle curve represents the mean relation derived from twenty simulations, the two outer curves illustrate the extreme results found in the twenty simulations. Smooth curves have been drawn through the data points for clarity.

been discussed - Chapter 2. The only reliable data at this scale comes from SFEM, who also find the turnover, and identify this portion of the covariance function with the power-law region observed in shallower samples, this conclusion appears to rely mainly on the lack of any significant clustering at larger scales.

The detection of significant large scale ($\theta > 0.2$ degrees) clustering in the deep samples described here leads to an alternative explanation for the small scale behaviour. The power-law behaviour seen in shallower samples is identified with the large scale clustering extending to scales of order 0.5 degrees. The very small scale turnover also occurs in the SP78 model simulation on occasions, as there is a deficiency of galaxy separations at scales $< 0.1h^{-1}$Mpc. The detailed behaviour varies from simulation to simulation, and is sensitive to the number of very large clusters which provide many small scale separations to the statistic. Again this is in good agreement with the observations. A test to ensure the observed turnover was not a product of the "merging" of close pairs of galaxies in the data was carried out by artificially merging images at larger scales. The change in the form of the covariance function was in the correct sense, but slope changes of 0.05 at most were found. Physically the reduction in the number of pairs at small separations must correspond to some additional physical process operating at small scales. The most obvious mechanism is dynamical friction (e.g. Ostriker and Turner 1979), the detailed effects depend on assumptions concerning the masses and extents of galactic halos but significant modification of the number of galaxy pairs on scales of order $0.1h^{-1}$Mpc can be expected under reasonable assumptions.

The case the covariance function observations described here are in excellent agreement with the predictions of the gravitational clustering

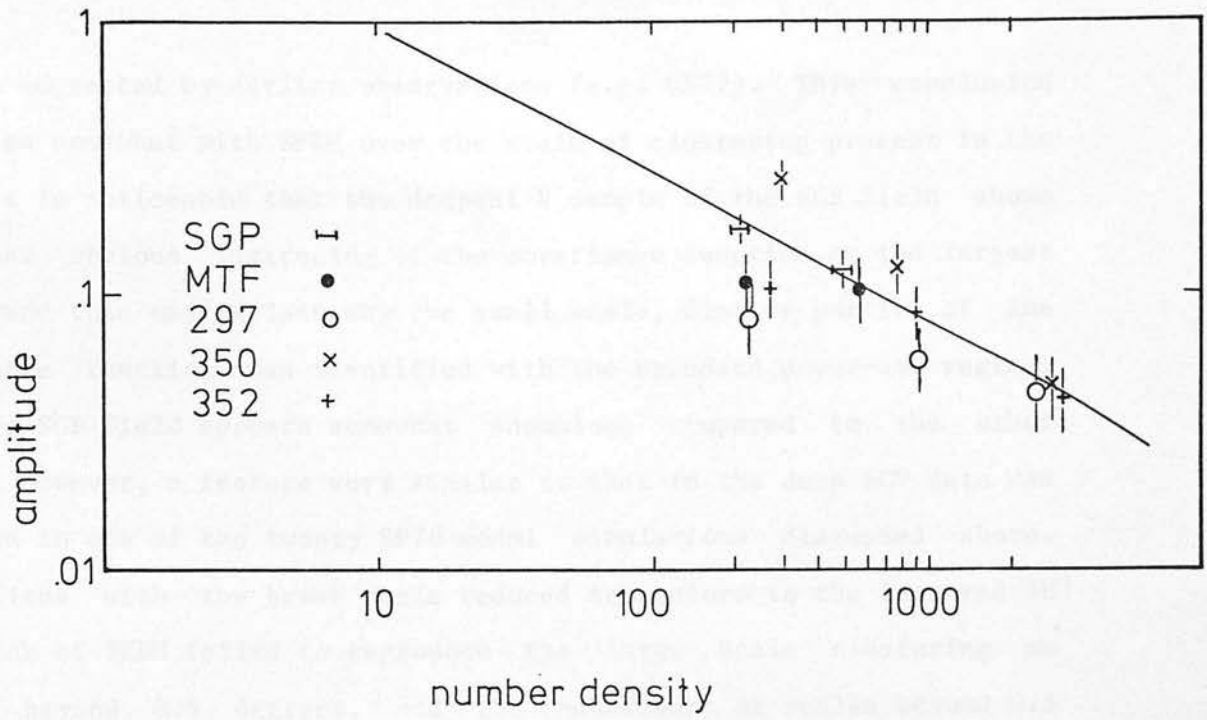


Fig. 9.6 The amplitude of the covariance function for all the samples evaluated at a scale of $0.5h^{-1}$Mpc. A scale chosen to avoid the highly variable behaviour between fields at larger scales. The predicted scaling behaviour for the standard spatial model is shown (also evaluated at $0.5h^{-1}$Mpc). Error bars shown do not take account of the large field to field variations found in the simulations.

picture suggested by earlier observations (e.g. GP77). This conclusion disagrees somewhat with SFEM over the scale of clustering present in the data. It is noticeable that the deepest V sample of the SGP field shows the most obvious flattening of the covariance function at the largest scale, and this may explain why the small scale, flatter portion of the covariance function was identified with the standard power-law regime. The deep SGP field appears somewhat anomalous compared to the other fields, however, a feature very similar to that in the deep SGP data was produced in one of the twenty SP78 model simulations discussed above. Simulations with the break scale reduced to conform to the favoured $3h^{-1}$ Mpc break of SFEM failed to reproduce the large scale clustering on scales beyond 0.4 degrees, and the behaviour at scales beyond 0.5 degrees was far more consistent from simulation to simulation than in the observational data.

The constraints that could be placed on the nature of large scale structure - the occurrence of filaments, sheets and their density contrasts for example - were disappointing. In the case of bubbles and sheets individual structures are not resolved; the structures typically subtend angles much larger than the field size of the observations. In the simulations, galaxy clustering follows standard models within the large scale structures, and the current observations are very insensitive to the properties of these structures. Galaxy overdensities in the range 2-20, with sheet major axis dimensions $20\text{-}50h^{-1}$ Mpc and bubble radii $15\text{-}25h^{-1}$ Mpc all gave 2D projections that could not be distinguished from the data. The amplitude and form of clustering being adjusted by modifying the small scale galaxy distribution within the structures.

Rather more can be said about filaments; with two small dimensions (~ $5h^{-1}$ Mpc) the individual structures produce identifiable features in the single plate fields. A uniform density filament results in an almost flat covariance function at scales smaller than the filament width, with a power-law drop of slope -1 on scales greater than the width but less than the filament length - the 2D covariance function averages over rings whose area increases with radius while the number of galaxy pairs remains constant with separation. Filament width is defined as the diameter containing 50% of the total galaxies associated with the structure. Detailed simulations confirmed the expected behaviour, the filament width is particularly important, as this determines the extent of the region where the probability of finding a galaxy about another galaxy is constant. This produces a covariance function with very different behaviour from that observed in the shallower samples, as well as those described here. The three-point function for the filament models was in close agreement with observations provided the filament width was not reduced below about $3h^{-1}$Mpc. A reduction below this scale allows the three-point function in a single field to pick up the shape dependence - the calculation of the three-point function extending to $5h^{-1}$ Mpc. The data show no strong shape dependence, although data and simulation results are noisy. For a reasonable range of filament parameters the fluctuations at large scale (0.5-2.0 degrees) were also within the range observed in the data. The simulations are not sensitive to the filament length because of the small sample size. Fig. 9.8 shows the allowed range of filament overdensity and width for the calculated two and three-point correlation functions to be consistent with the data. The filament length was $50h^{-1}$Mpc. A wide range of filament models are able to fit the data, including models which fit the nearby redshift surveys well. The fact that a standard heirarchical model as well as relatively extreme filamentary models can be made to fit the data

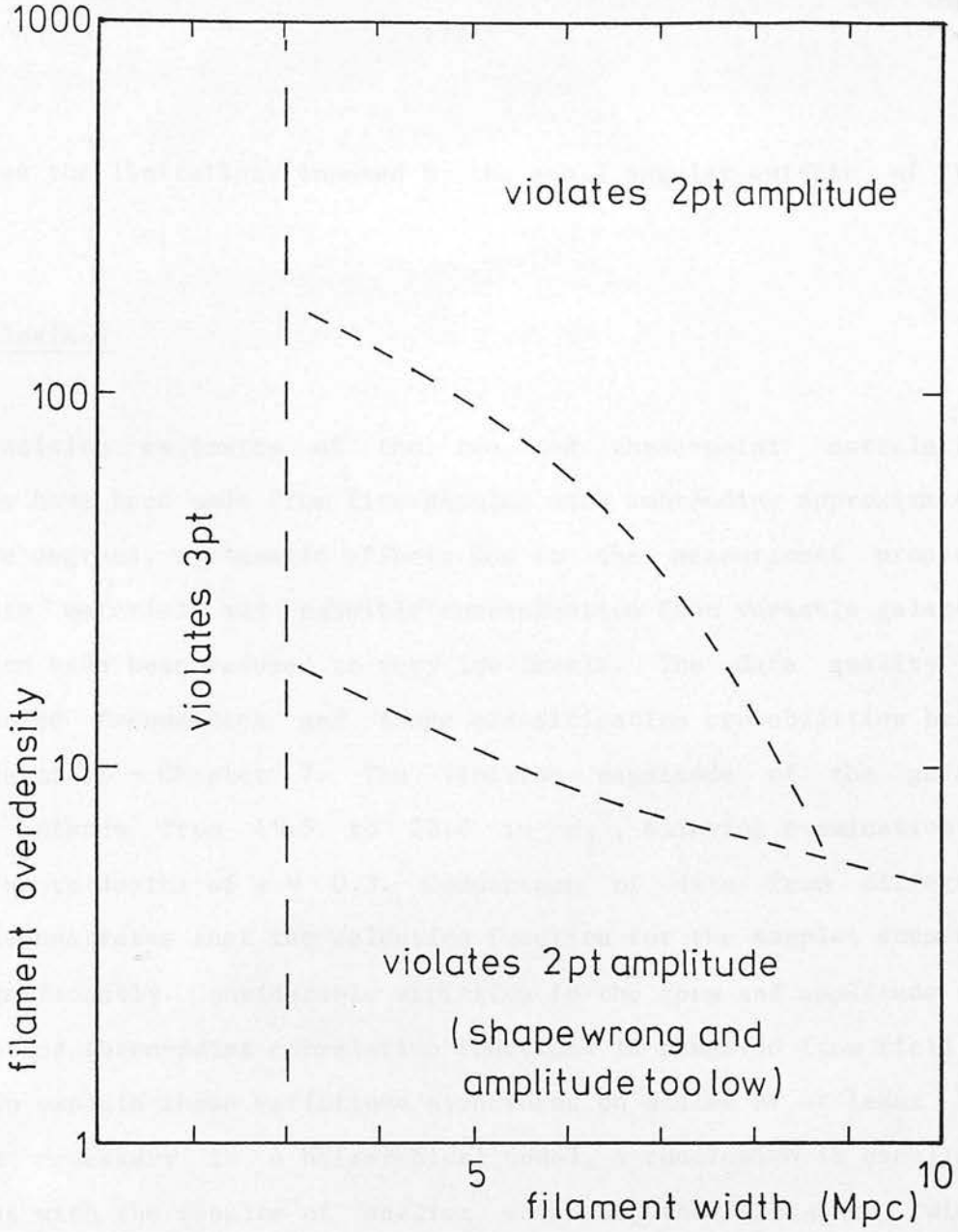


Fig. 9.8 The region of overdensity-structure width for linear filaments that can be made consistent with the data of this Chapter by manipulation of the small scale ($\lesssim 5h^{-1}$ Mpc) clustering. Very high density, wide filaments are ruled out by the amplitude and slope of the covariance function. Very narrow filaments are detectable with the three-point function estimates in a single field. The filament length is $50h^{-1}$Mpc but the results from the small fields are insensitive to this scale within a factor two.

emphasises the limitations imposed by the small angular extents of the samples.

9.5 Conclusions

High precision estimates of the two and three-point correlation functions have been made from five samples each subtending approximately 21 square degrees, systematic effects due to the measurement process, the plate material and possible contamination from variable galactic extinction have been reduced to very low levels. The data quality is high; noise frequencies and image classification probabilities being well determined - Chapter 7. The limiting magnitude of the galaxy samples extends from 19.5 to 22.0 in m_J, allowing examination of clustering to depths of $z = 0.3$. Comparison of data from different plates demonstrates that the selection function for the samples does not vary significantly. Considerable variation in the form and amplitude of the two and three-point correlation functions is observed from field to field. To explain these variations structures on scales of at least $10h^{-1}$ Mpc are necessary in a heirarchical model, a conclusion in excellent agreement with the results of earlier work on the Lick and Zwicky catalogues (e.g Peebles 1980). Alternatively models with very large scale filamentary structures containing galaxies distributed according to heirarchical or power-law models on small scales also fit the observational data. The filamentary structure conforms well to the predictions of the adiabatic fluctuation galaxy formation picture. The existence of large scale structure on scales of $10h^{-1}$ Mpc is confirmed but much larger samples are required to discriminate between the possible forms of this structure.

References

Aarseth,S.J.& Saslaw,W.C.,1982.Astrophys. J.,258,L7.

Abadi,H.I.& Edmunds,M.G.,1978.Astron. Astrophys.,70,189.

Abell,G.O.,1958.Astrophys. J. Supp.,3,211.

Abell,G.O.,1961.Astron. J.,66,607.

Allen,C.W.,1973. "Astrophysical Quantities", Athlone Press, London.

Appenzeler,I.,1975.Astron. Astrophys.,38,113.

Appleton,P.N.& Davis,R.D.,1982.M.N.R.A.S., In press.

Bahcall,N.A.,1977.Ann. Rev. Astron. Astrophys.,15,505.

Bahcall,J.N.& Soneira,R.M.,1981.Astrophys. J.,246,122.

Bahcall,J.N.& Soneira,R.M.,1983.Astrophys. J., In press.

Baker,P.& Burton,W.B.,1975.Astrophys. J.,198,281.

Balzano,V.A.& Weedman,D.W.,1982.Astrophys. J.,255,L1.

Bevington,P.R.,1969. "Data Reduction and Error Analysis for the Physical Sciences", McGraw Hill, New York.

Bonometto,S.A.& Lucchin,F.,1980.Astron. Astrophys.,82,287.

Bonometto,S.A.&. Sharp,N.A.,1980.Astron. Astrophys.,92,222.

Bothun,G.D.& Schommer,R.A.,1982.Astrophys. J.,255,L23.

Boughn,S.P.,Cheng,E.S.& Wilkerson,D.T.,1981.Astrophys. J.,243,L113.

Brown,G.S.,1978. Publ. Univ. Texas., No. 11.

Bruzual A.,G.,1981. Ph.D. Thesis, University of California, Berkeley.

Burstein,D.& Heiles,C.,1978.Astrophys. J.,225,40.

Burstein,D.& Heiles,C.,1982.Astron. J.,87,1165.

Campbell,A.W.,1981.M.Sc. Thesis, University of Edinburgh.

Cannon,R.D.,Hawarden,T.G.,Sim,M.E.& Tritton,S.B.,1978.Occ. Rep. Roy. Obs. Edin.,4.

Carter,D.,1980.M.N.R.A.S.,190,307.

Carter,D.& Godwin,J.G.,1979.M.N.R.A.S.,187,711.

Chamaraux,P.,Balkowski,C.& Gerard,E.,1980.Astron. Astrophys.,83,38.

Chincarini,G.,Rood,H.J.& Thompson,L.A.,1981.Astrophys. J.,249,L47.

Coleman,G.D.,Wu,C-C.& Weedman,D.W.,1980.Astrophys. J. Supp.,43,393.

Cooke,J.A.,Emerson,D.,Kelly,B.D.,MacGillivray,H.T.& Dodd,R.J.,1981.
M.N.R.A.S.,196,397.

Corwin,H.G.,1981.Ph.D. Thesis, University of Edinburgh.

Dautcourt,G.,Kempe,K.,Richter,L.& Richter,N.,1978.
Astron. Nachr.,299,177.

Davis,M.,1980. in "Objects of High Redshift", p57, eds. G.O. Abell &
P.J.E. Peebles, D. Reidel, Dordrecht.

Davis,M.,Groth,.E.J.& Peebles,P.J.E.,1977.Astrophys. J.,212,L107.

Davis,M.& Huchra,J.P.,1982.Astrophys. J.,254,437.

Davis,M.,Huchra,J.P.,Latham,D.& Tonry,J.,1982.Astrophys. J.,253,423.

Davis,M.,Tonry,J.,Huchra,J.& Latham,D.W.,1980.Astrophys. J.,238,L113.

Davis,M.& Wilkerson,D.T.,1974.Astrophys. J.,192,251.

Dawe,J.A.,1981.UKSTU Internal Rep., SSO/01/81.

Dawe,J.A.& Metcalfe,N.,1981.UKSTU Internal Rep., SSO/02/81.

Dawe,J.A.& Metcalfe,N.,1982.Proc. Astron. Soc. Australia, In press.

Dickey,J.M.,Weisberg,J.M.,Rankin,J.M.& Boriakoff,V.,1981.
Astron. Astrophys.,101,332.

Dodd,R.J.& MacGillivray,H.T.,1980. in "ESO Workshop on Two Dimensional
Photometry", p391, eds. P. Craine & K. Kjar, ESO.

Doroshkevich,A.G.,Zeldovich,Ya.B.& Syunyaev,R.A.,1978.
Sov. Astron.,22,523.

Doroshkevich,A.G.,Kotok,E.V.,Novikov,I.D.,Polyudov,A.N.,
Shandarin,S.F.& Sigov,Yu.S.,1980.M.N.R.A.S.,192,337.

Efstathiou,G.& Eastwood,J.W.,1981.M.N.R.A.S.,194,503.

Einasto,J.,Joeveer,M.& Saar,E.,1980.Nature,283,47.

Ellis,J.,1982.Phil. Trans. R. Soc. Lond. A,307,121.

Ellis,R.S.,1982. in "The Origin and Evolution of the Galaxies",
eds. B.J.T. Jones & J. Jones, D. Reidel, Dordrecht.

Ellis,R.S.,Fong,R.& Phillipps,S.,1977.M.N.R.A.S.,181,163.

Fabbri,R.,Guidi,I.,Melchiorri,F.& Natale,V.1980.Phys. Rev. Lett.,44, 1563.

Faber,S.M.& Gallagher,J.S.,1979.Ann. Rev. Astron. Astrophys.,17,135.

Fall,S.M.,1979.Rev. Mod. Phys.,51,21.

Fall,S.M.& Tremaine,S.,1977.Astrophys. J.,216,682.

Felton,J.E.,1977.Astron. J.,82,861.

Fesenko,B.I.,1978.Sov. Astron. Lett.,4,288.

Fesenko,B.I.,1979.Sov. Astron.,23,524.

Fisher,J.R.& Tully,R.B.,1981.Astrophys. J. Supp.,47,139.

Fry,J.N.& Peebles,P.J.E.,1978.Astrophys. J.,221,19.

Fry,J.N.& Peebles,P.J.E.,1980.Astrophys. J.,238,785.

Fry,J.N.& Seldner,M.,1982.Astrophys. J.,259,474.

Godwin,J.G.,Metcalfe,N.,& Peach,J.V.,1982.M.N.R.A.S.,In press.

Gott,J.R.,1982.Nature,295,304.

Gott,J.R.& Turner,E.L.,1976.Astrophys. J.,209,1.

Gott,J.R.& Turner,E.L.,1979.Astrophys. J.,232,L79.

Gregory,S.A.& Thompson,L.A.,1978.Astrophys. J.,222,784.

Gregory,S.A.,Thompson,L.A.& Tifft,W.G.,1981.Astrophys. J.,243,411.

Gregory,S.A.,Chincarini,G.,Rood,H.J.& Thompson,L.A.,1978. Astrophys. J.,253,724.

Groth,E.J.,1980. in "Objects of High Redshift", p31, eds. G.O. Abell & P.J.E. Peebles, D. Reidel, Dordrecht.

Groth,E.J.& Peebles,P.J.E.,1977.Astrophys. J.,217,385.

Gunn,J.E.,Longair,M.S.& Rees,M.J.,1978. "Observational Cosmology", Geneva Observatory, Sauvernay.

Guth,A.,1981.Phys. Rev. D,23,347.

Hauser,M.G.& Peebles,P.J.E.,1973.Astrophys. J.,185,757

Hawkins,M.R.S.,1981.M.N.R.A.S.,194,1013.

Heiles,C.,1980.Astrophys. J.,235,833.

Heiles,C.,Stark,A.A.& Kulkarni,S.,1981.Astrophys. J.,247,L73.

Hewett,P.C.,1982.M.N.R.A.S.,201,867.

Hilditch,R.W.,Hill,G.& Barnes,J.V.,1976.M.N.R.A.S.,176,175.

Himmes,A.& Biermann,P.,1980.Astron. Astrophys.,86,11.

Holmberg,E.B.,1974.Astron. Astrophys.,35,21.

Hubble,E.,1934.Astrophys. J.,79,8.

Huchra,J.,1976.Astron. J.,81,952.

Huchra,J.P.& Geller,M.J.,1982.Astrophys. J.257,423

Jarvis,J.F.& Tyson,J.A.,1981.Astron. J.,86,476.

Joeveer,M.,Einasto,J.& Tago,E.,1978.M.N.R.A.S.,185,357.

Jones,B.T.,1976.Rev. Mod. Phys.,48,107.

Kalberla,P.M.W.,Mebold,U.& Reid,W.,1980a.Astron. Astrophys.,82,275.

Kalberla,P.M.W.,Mebold,U.& Velden,J.,1980b.Astron. Astrophys. Supp.,39,337.

Kirshner,R.P.,Oemler,A.& Schechter,P.L.,1978.Astron. J.,83,1549.

Kirshner,R.P.,Oemler,A.& Schechter,P.L.,1979.Astron. J.,84,951.

Kirshner,R.P.,Oemler,A.,Schechter,P.L.& Schectman,S.A.,1981.
Astrophys. J.,248,L57.

Knude,J.,1979.Astron. Astrophys.,71,344.

Knude,J.,1981.Astron. Astrophys.,98,74.

Koo,D.C.,& Kron,R.G.,1982.Astron. Astrophys.,105,107.

Kron,R.G.,1980.Astrophys. J. Supp.,43,305.

Lake,G.& Tremaine,S.,1980.Astrophys. J.,238,L13.

Latham,D.& Furenlid,I.,1976.AAS Photo-Bulletin,11,11.

Limber,D.N.,1953.Astrophys. J.,17,134.

Longair,M.S.& Seldner,M.,1979.M.N.R.A.S.,189,433.

Luyten,W.J.,1968.M.N.R.A.S.,139,221.

MacGillivray,H.T.,1976.Ph.D. Thesis, University of Edinburgh.

MacGillivray,H.T.& Dodd,R.J.,1979.M.N.R.A.S.,186,69.

MacGillivray,H.T.& Dodd,R.J.1980.M.N.R.A.S.,193,1.

McKee,C.F.& Ostriker,J.P.,1977.Astrophys. J.,218,148.

McNally,B.,1979.M.Sc. Thesis, Universiy of Edinburgh.

Masson,C.,1979.M.N.R.A.S.,188,261.

Moffat,A.F.J.,1969.Astron. Astrophys.,3,455.

Mukhanov,V.F.& Shcherbanovskii,A.L.,1979.Sov. Astron. Lett.,5,311.

Neyman,J.,Scott,E.L.& Shane,C.D.,1953.Astrophys. J.,117,92.

Nilson,P.,1973.Uppsala astr. Obs. Ann.,6.

Noonan,T.W.,1973.Astron. J.,78,227.

Okamura,S.,Davenhall,A.C.& MacGillivray,H.T.,1983.Astron. Astrophys., In press.

Oort,J.H.,1983.Ann. Rev. Astron. Astrophys.,21,In press.

Osmer,P.,1981.Astrophys. J.,247,762.

Ostriker,J.P.,Peebles,P.J.E.& Yahil,A.,1974.Astrophys. J.,193,L1.

Ostriker,J.P.& Turner,E.L.,1979.Astrophys. J.,234,785.

Partridge,R.B.,1974.Astrophys. J.,192,241.

Peacock,J.A.& Gull,S.F.,1981.M.N.R.A.S.,196,611.

Peebles,P.J.E.,1973.Astrophys. J.,185,413.

Peebles,P.J.E.,1974.Astron. Astrophys.,32,197.

Peebles,P.J.E.,1975.Astrophys. J.,196,647.

Peebles,P.J.E.,1979.M.N.R.A.S.,189,89.

Peebles,P.J.E.,1980. "The Large Scale Structure of the Universe", Princeton University Press, Princeton.

Peebles,P.J.E.,1981.Astrophys. J.,243,L119.

Peebles,P.J.E.& Groth,E.J.,1975.Astrophys. J.,196,1.

Peebles,P.J.E.& Hauser,M.G.,1974.Astrophys. J. Supp.,28,19.

Pence,W.,1976.Astrophys. J.,203,39.

Penzias,A.A.& Wilson,R.W.,1965.Astrophys. J.,142,419.

Peterson,B.A.,Ellis,R.S.,Kibblewhite,E.J.,Bridgeland,M.T.,Hooley,T.& Horne,D.,1979.Astrophys. J.,233,L109.

Phillipps,S.,Ellis,R.S.& Strong,A.W.,1981.M.N.R.A.S.,197,151.

Phillipps,S.,Fong,R.,Ellis,R.S.,Fall,S.M.& MacGillivray,H.T.,1978.

M.N.R.A.S.,182,673.

Pickup,G.E.,1979.PhD Thesis, University of Edinburgh.

Press,W.H.& Davis,M.,1982.Astrophys. J.,259,449.

Press,W.H.& Lightman,A.P.,1978.Astrophys. J.,219,L73.

Prichet,C.,& Kline,M.I.,1981.Astron. J.,86,1859.

Reaves,G.,1974.Sov. Astron.,18,307.

Rees,M.,1980. in "Objects of High Redshift", p207, eds. G.O. Abell & P.J.E. Peebles, D. Reidel, Dordrecht.

Reid,I.,R.& Gilmore,G.,1982.M.N.R.A.S.,201,68.

Rood,H.J.1981.Rep. Prog. Phys.,44,1077.

Rudnicki,K.,Dworak,T.Z.,Flin,P.,Baranowski,B.& Sendrolowski,A.,1973. Acta Cosmol.,1,7.

Sanduleak,N.& Pesch,P.,1982.Astrophys. J.,258,L11.

Sargent,W.L.W.,Young,P.J.,Boksenberg,A.& Tytler,D.,1980. Astrophys. J. Supp.,42,41.

Savage,B.D.& Mathis,J.S.,1979.Ann. Rev. Astron. Astrophys.,17,73.

Sebok,W.L.,1979.Astron. J.,84,1526.

Seldner,M.,& Peebles,P.J.E.,1977.Astrophys. J.,215,703.

Seldner,M.,& Peebles,P.J.E.,1981.M.N.R.A.S.,194,251.

Seldner,M.,Siebers,B.,Groth,E.J.& Peebles,P.J.E.,1977. Astron. J.,82,249.

Shane,C.D.& Wirtanen,C.A.,1967.Publ. Lick Obs.,22,1

Shanks,T.,1979.M.N.R.A.S.,186,583.

Shanks,T.,1982. in "Progress in Cosmology", p335, ed. A.W. Wolfendale, D. Reidel, Dordrecht.

Shanks,T.,Fong,R.,Ellis,R.S.& MacGillivray,H.T.,1980a. M.N.R.A.S.,192,209.

Shanks,T.,Phillipps,S.& Fong,R.,1980b.M.N.R.A.S.,191,47p.

Shapley,H.,1940.Proc. Nat. Acad. Sci.,26,599.

Sharp,N.A.,1979.Astron. Astrophys.,74,312.

Sharp,N.A.,1981.M.N.R.A.S.,195,857.

Soneira,R.M.& Peebles,P.J.E.,1978.Astron. J.,83,845.

Spitzer,L.,1978. "Physical Processes in the Interstellar Medium", p156, John Wiley and Sons Inc., New York.

Stobie,R.S.,1980.J.B.I.S.,33,323.

Stobie,R.S.,Smith,G.M.,Lutz,R.K.& Martin,R.,1979. "Image Processing in Astronomy", p48, eds. G. Sedmak, M. Capaccioli & R.J. Allen, Osservatorio Astronomico di Trieste.

Strong,A.W.& Lebrun,F.,1982.Astron. Astrophys.,105,159.

Tarenghi,M.,Chincarini,G.,Rood,H.J.& Thompson,L.A.,1980. Astrophys. J.,235,7.

Tarenghi,M.,Tifft,W.G.,Chincarini,G.,Rood,H.J.& Thompson,L.A.,1979. Astronphys. J.,234,793.

Totsuji,H.& Kihara,T.,1969.P.A.S.J.,21,221.

Tritton,K.P.,1981. in "Astronomical Photography 1981", p1, eds. J.-L. Heudier & M.E. Sim, Centre National Recherche Scientifique, Nice.

Tully,R.B.,1980.Astrophys. J.,237,390.

Tully,R.B.,1982,Astrophys. J.,257,389.

Turner,E.L.,1976.Astrophys. J.,208,20.

Turner,E.L.& Gott,J.R.,1976.Astrophys. J.,209,6.

Tyson,J.A.,& Jarvis,J.F.,1980. in "Objects of High Redshift", p1, eds. G.O. Abell & P.J.E. Peebles, D. Reidel, Dordrecht.

de Vaucouleurs,G.1958.Applied Optics,1,1513.

de Vaucouleurs,G.,1971.P.A.S.P.,83,113.

UKSTU,1981.Newsletter,3,11.

Wall,J.V.,1980.Phil. Trans. R. Soc. Lond. A,296,367.

Wallace,P.T.& Tritton,K.P.,1979.M.N.R.A.S.,189,115.

Weilen,R.,1974. in "Highlights in Astronomy", 3, p395, ed. G. Contopoulos, D. Reidel, Dordtrecht.

Weinberg,S.,1972. "Gravitation and Cosmology", John Wiley and Sons Inc,

New York.

White,S.D.M.,1979.M.N.R.A.S.,186,145.

White,S.D.M.& Valdes,F.,1980.M.N.R.A.S.,190,55.

Wilkerson,D.T.,1982.Phil. Trans. R. Soc. Lond. A,307,55.

Williams,B.A.& Kerr,F.J.,1981.Astron. J.,86,953.

Yahil,A.,Sandage,A.& Tammann,G.A.,1980.Astrophys. J.242,448.

Zeldovich,Ya.B.,1978. in "The Large Scale Structure of the Universe", p409, eds. M.S. Longair & J. Einasto, D. Reidel, Dordrecht.

Zwicky,F.,Herzog,E.,Wild,P.,Karpowicz,M.& Kowal,C.T.,1961-1968."Catalogue of Galaxies and Clusters of Galaxies", in 6 vols., California Institute of Technology, Pasadena.

Appendix 1

A number of diagrams showing the distribution of images in COSMOS parameter spaces are included here - plots of this type with well determined image classifications are not currently available. The two image samples are described in Chapter 5. Symbol types are maintained throughout and defined in Fig. 5.1.

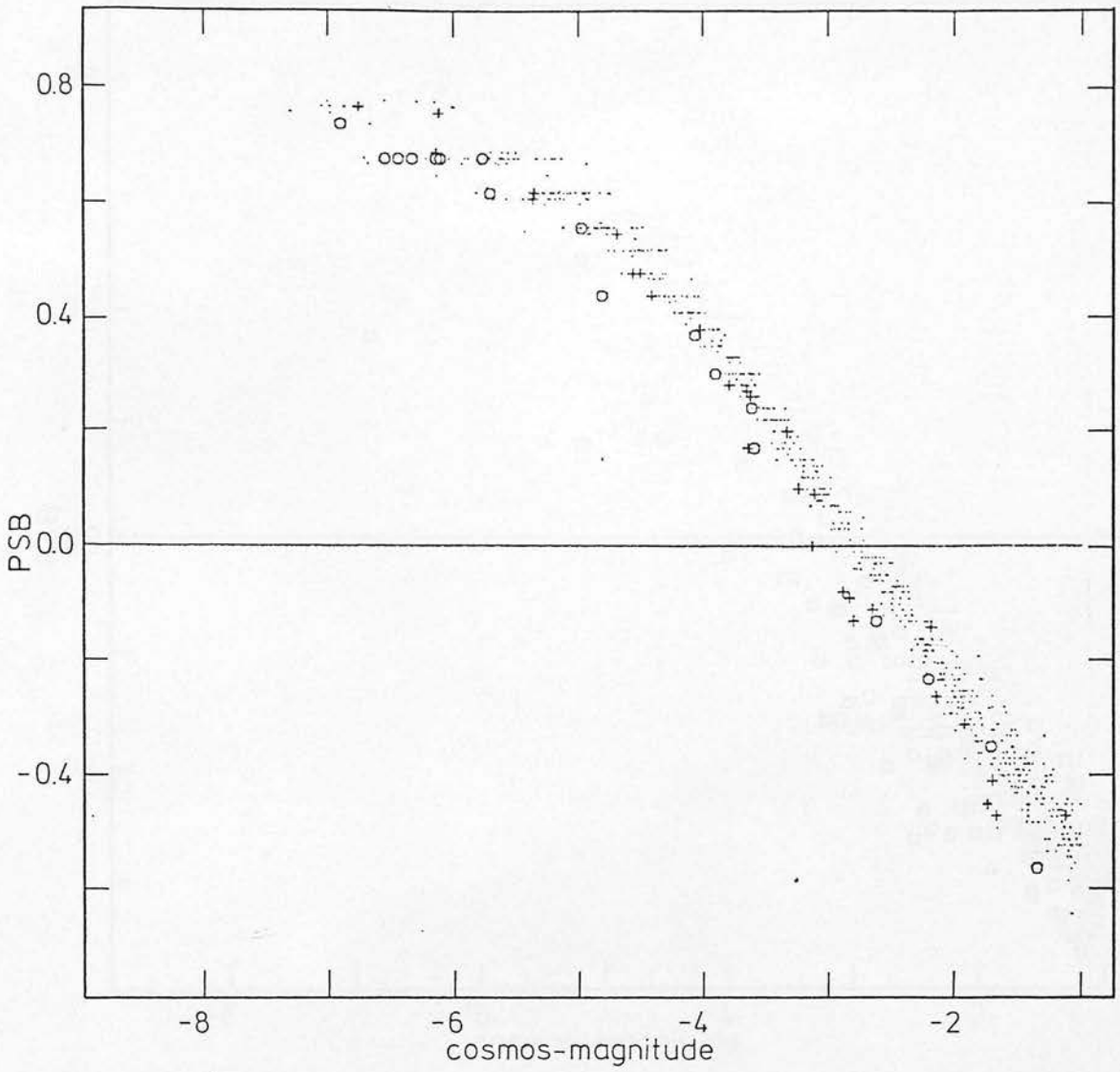


Fig. A1.1 The distribution of 1100 images from plate V3475 in cosmos-magnitude vs. peak surface brightness space.

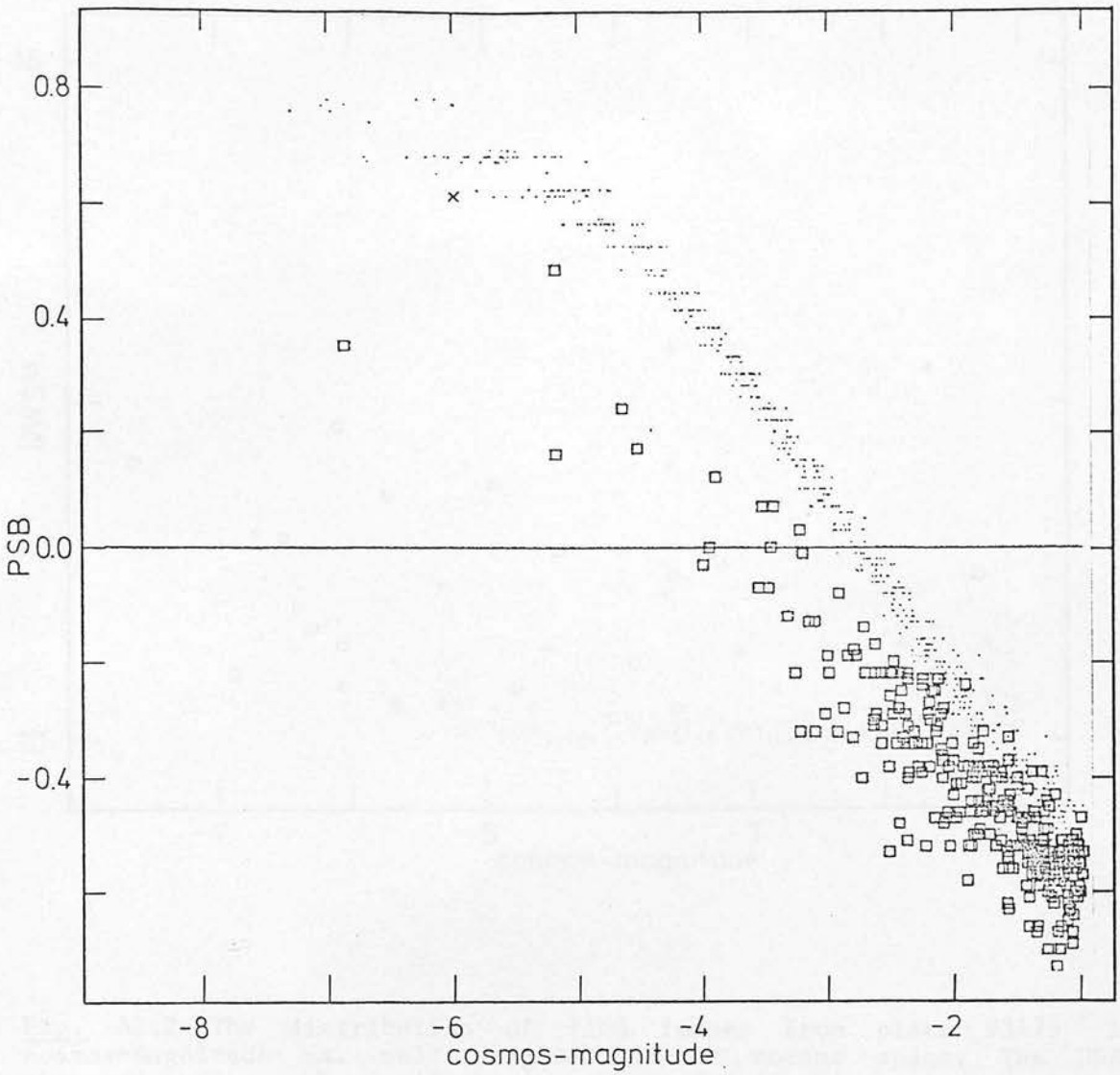


Fig. A1.1b

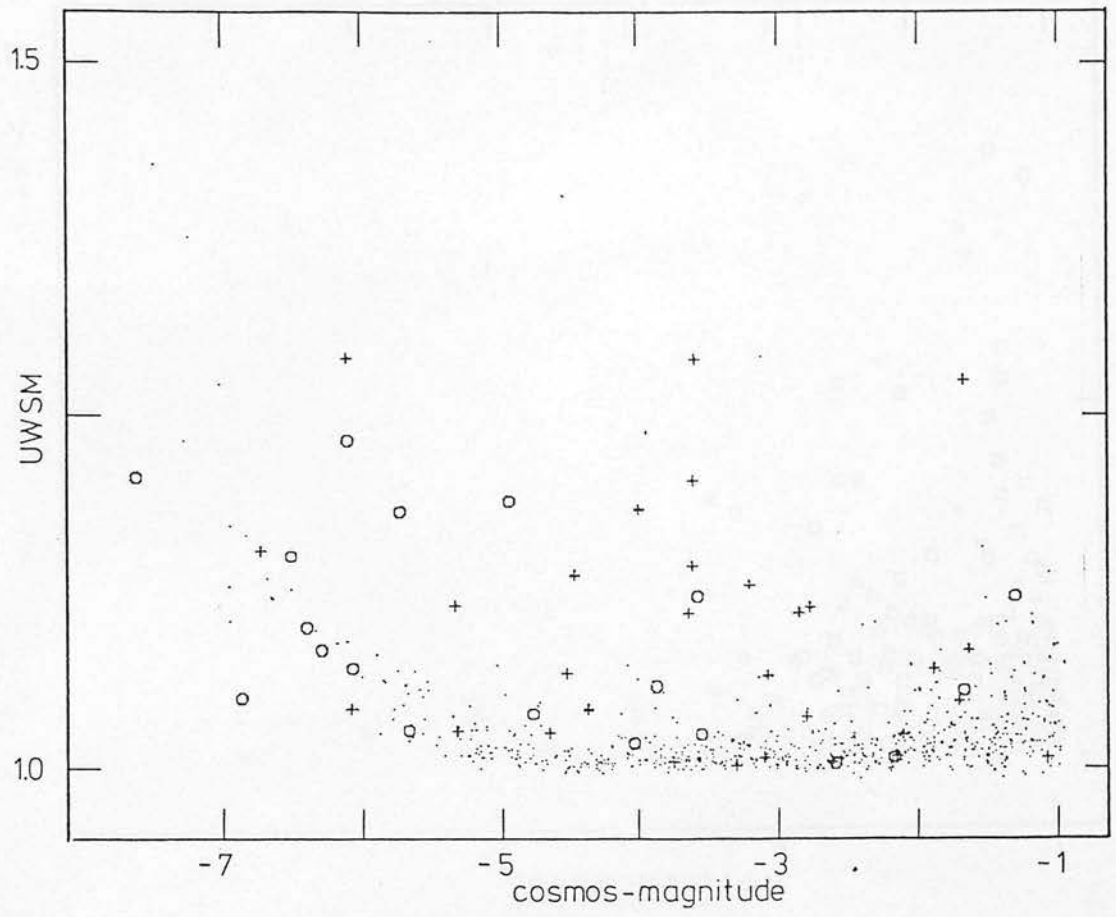


Fig. A1.2 The distribution of 1100 images from plate V3475 in cosmos-magnitude vs. unit weighted second moment space. The UWSM parameter measures how well the image limiting isophote is approximated by an ellipse; an elliptical isophote results in a value of one for the UWSM.

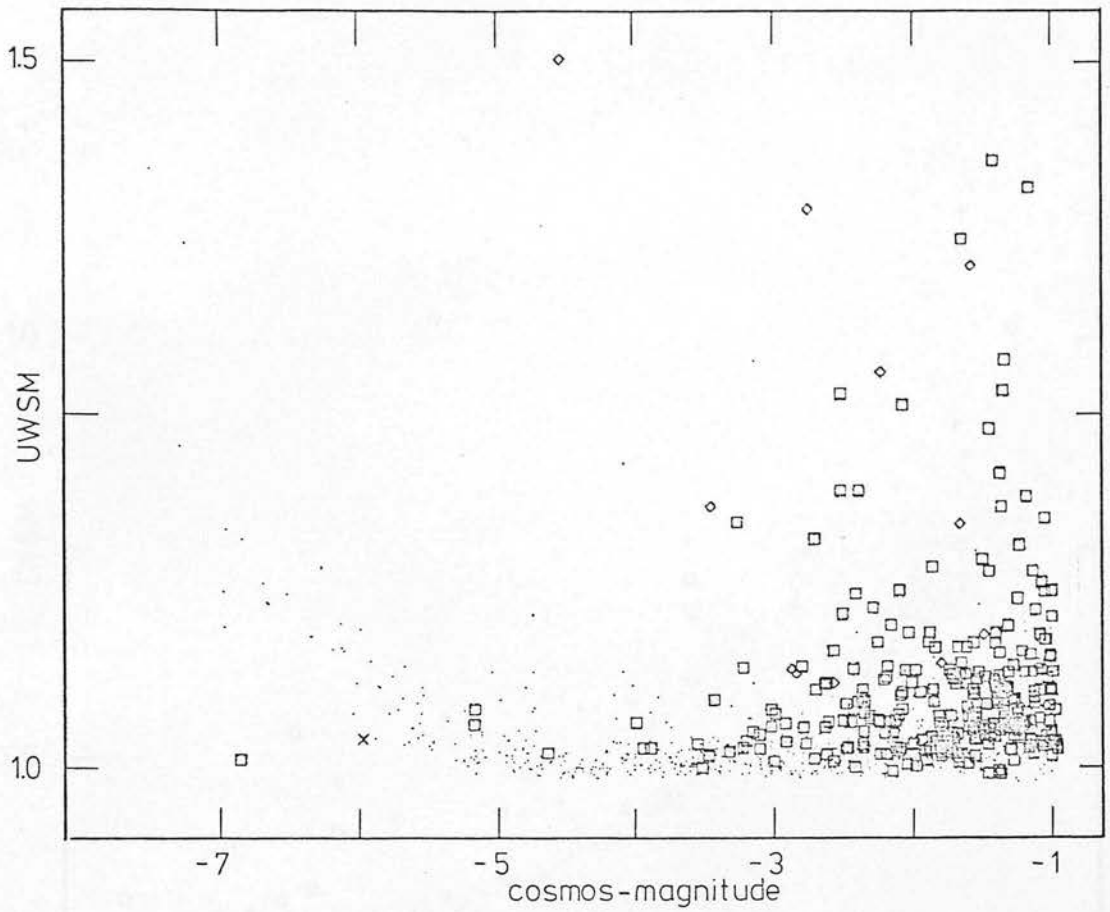


Fig. A1.2b

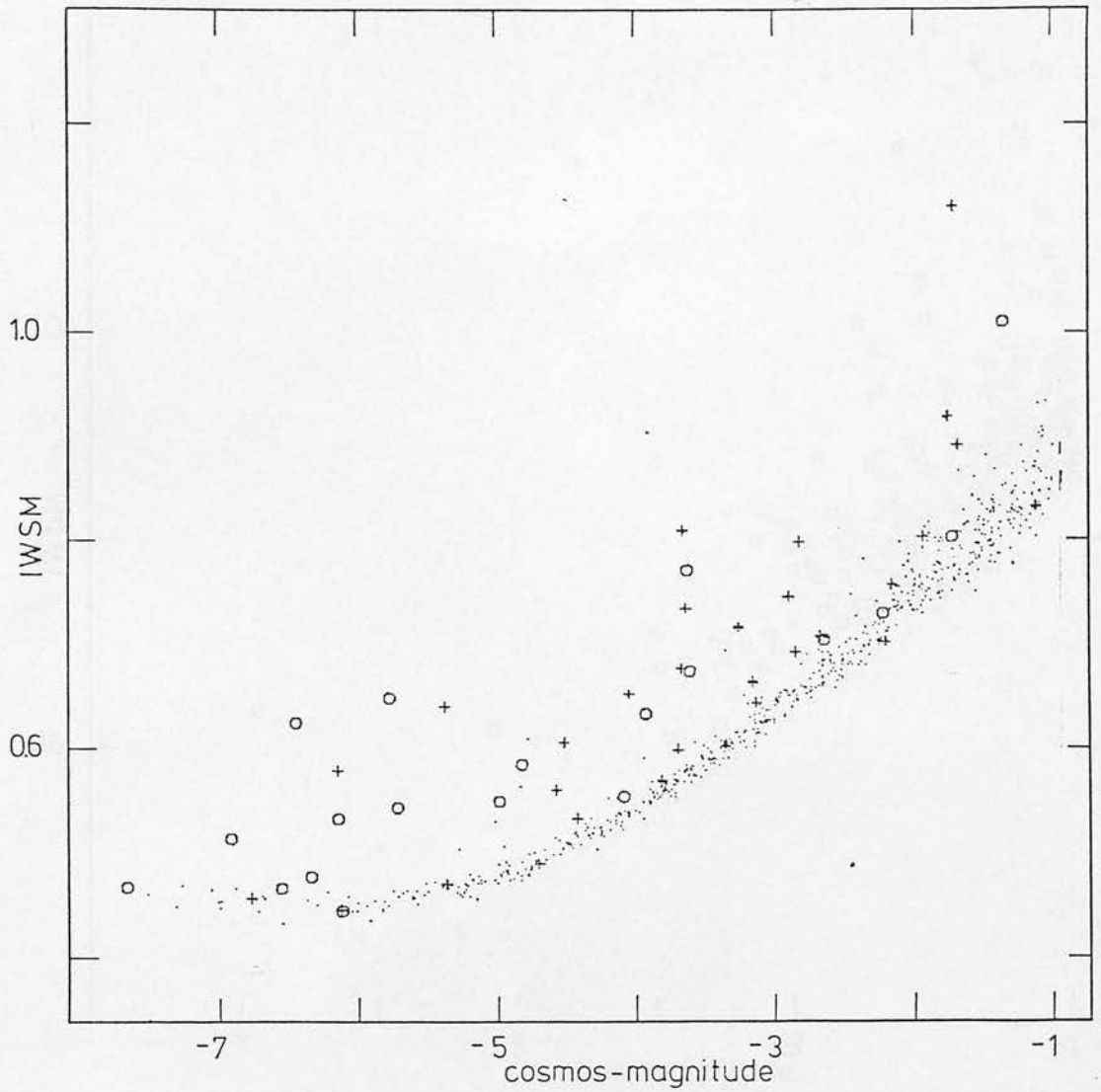


Fig. A1.3 The distribution of 1100 images from plate V3475 in cosmos-magnitude vs. intensity weighted second moment space. The IWSM parameter is analagous to a moment of inertia - more diffuse images have a larger IWSM.

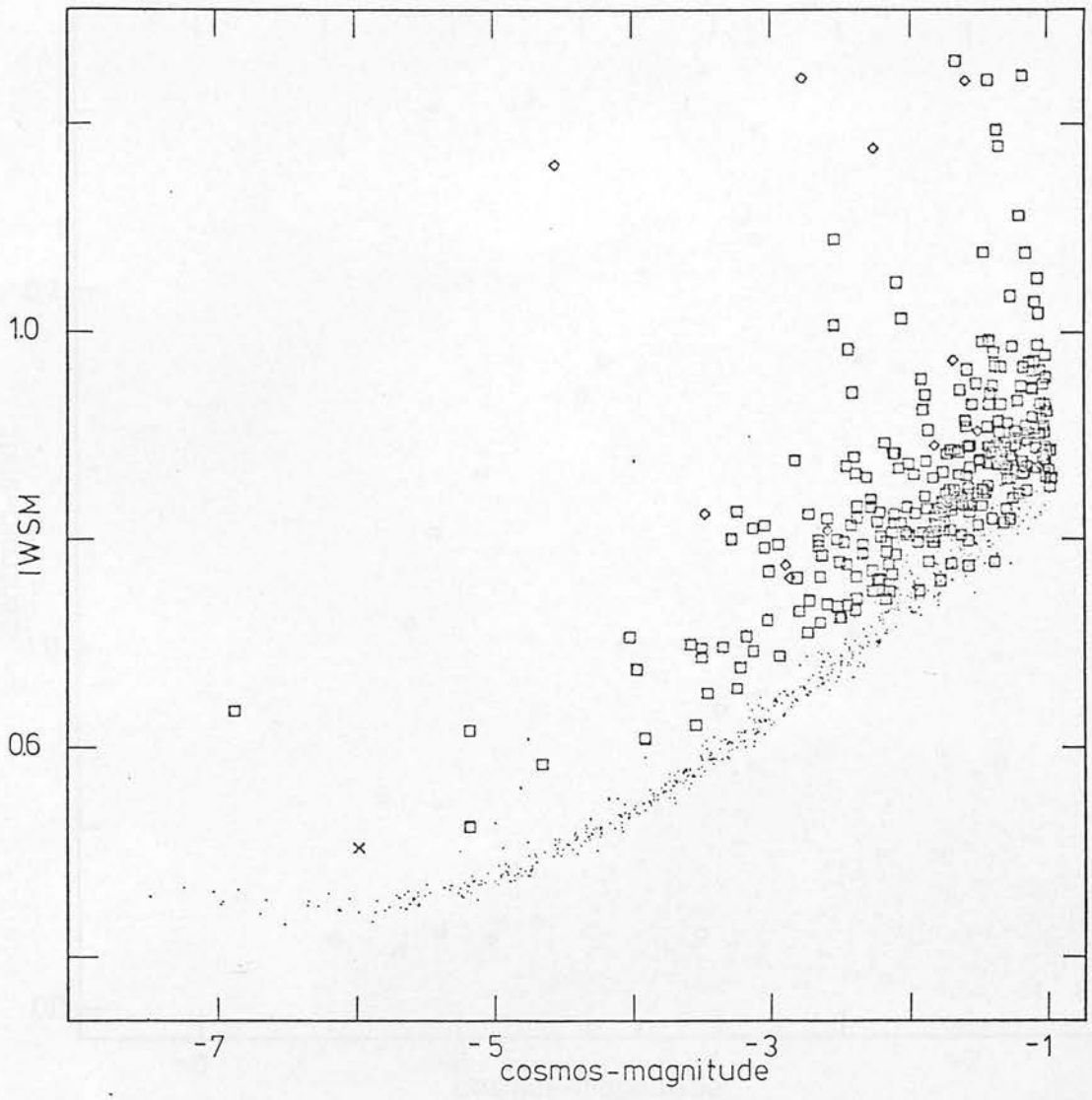


Fig. A1.3b

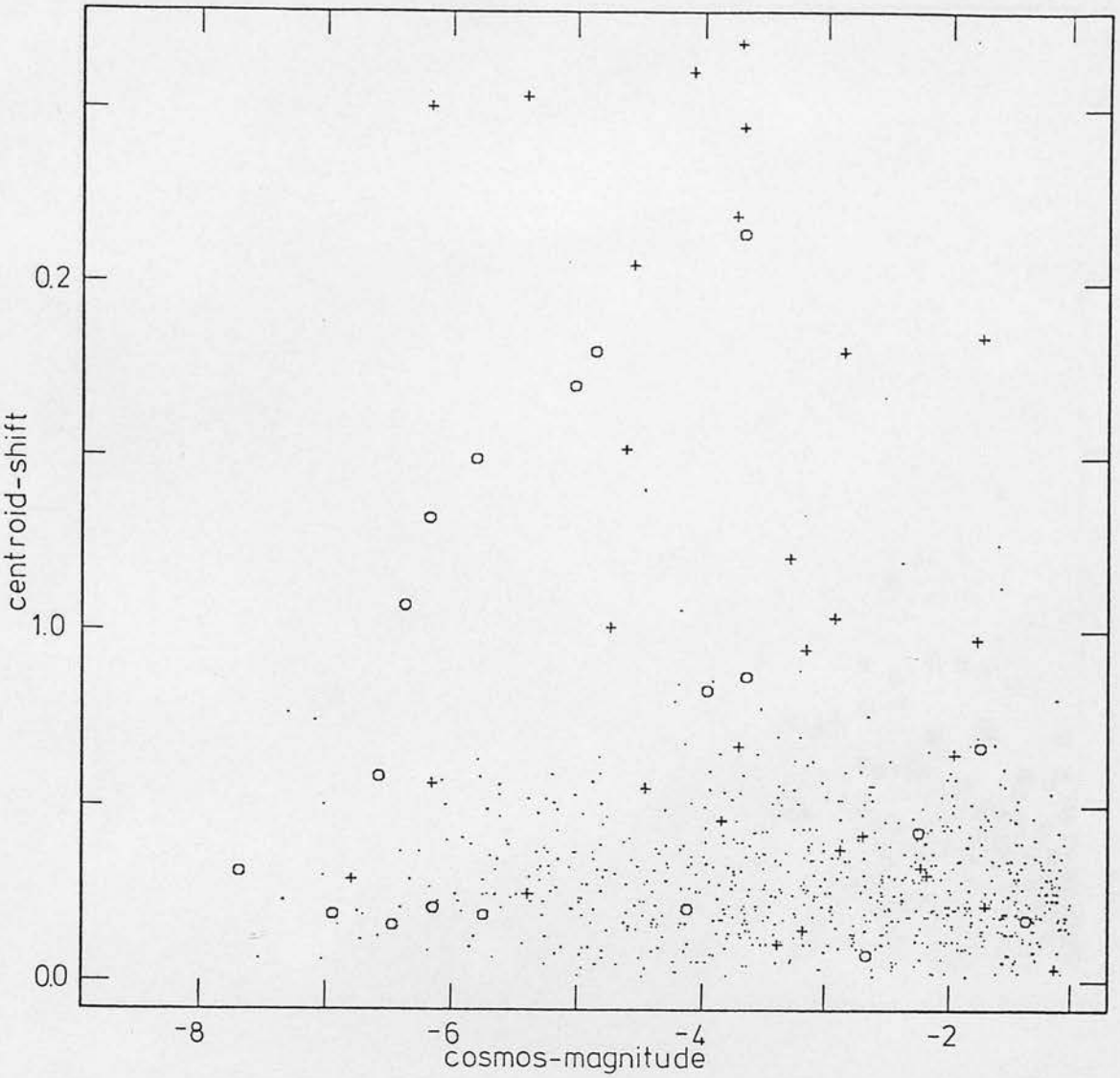


Fig. A1.4 The distribution of 1100 images from plate V3475 in cosmos-magnitude vs. relative image centroid shift. The parameter measures the fractional shift (in units of the image radius) between the unit weighted and intensity weighted image centroids. The parameter is large for images with strongly asymmetric profiles. Many merged images have large values of this parameter.

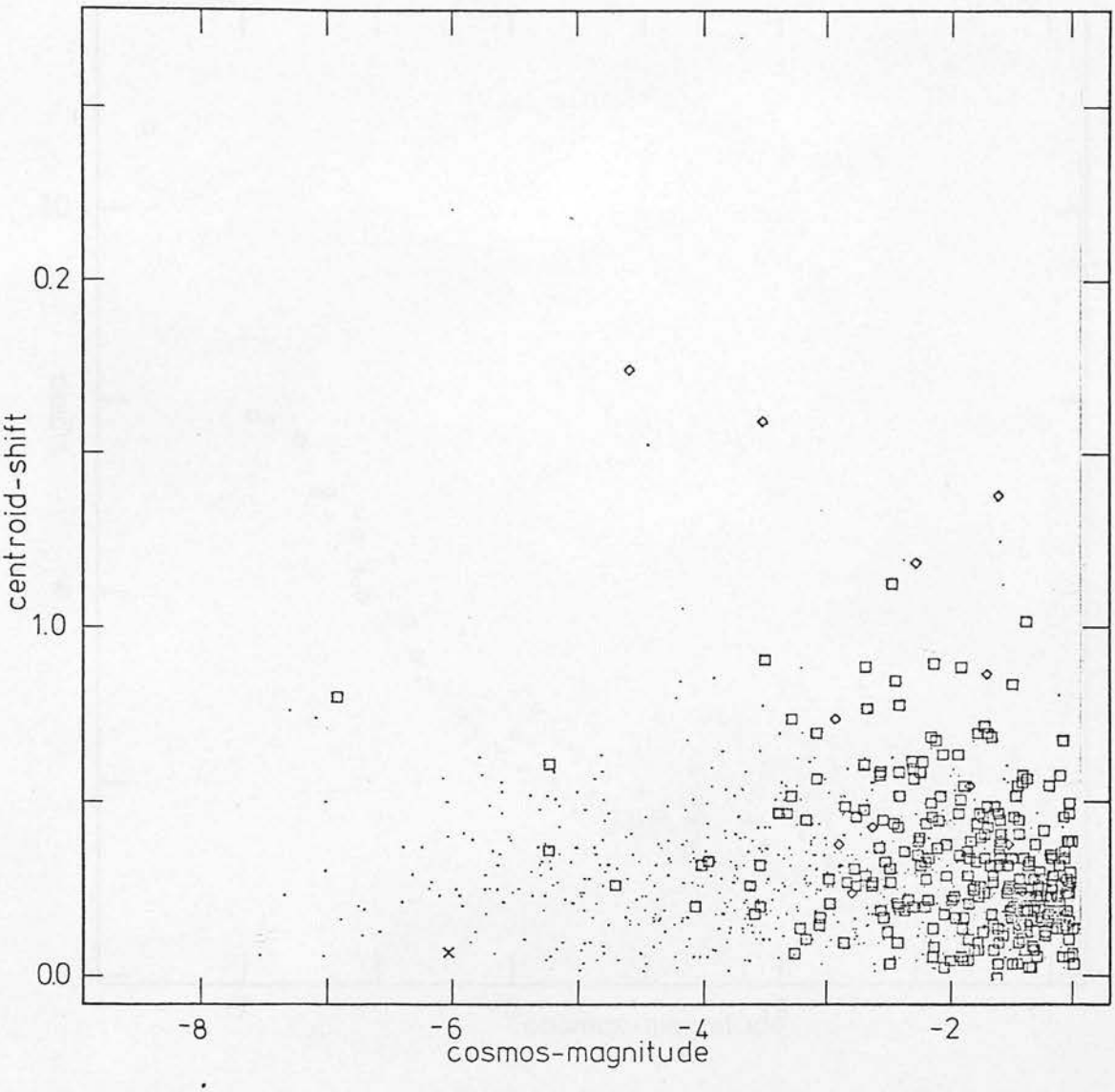


Fig. A1.4b

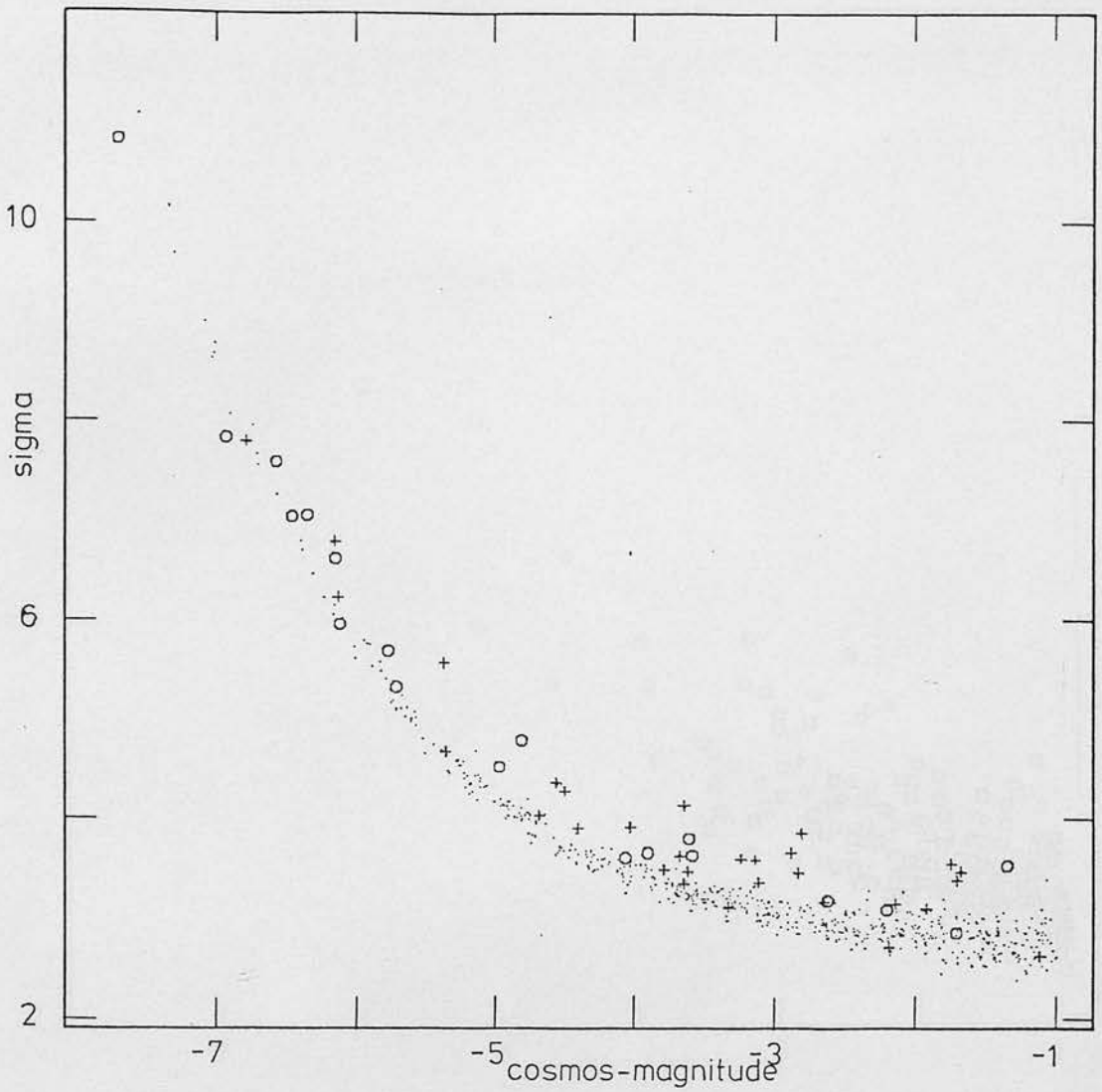


Fig. A1.5 The distribution of 1100 images from plate V3475 in cosmos-magnitude vs. equivelant gaussian dispersion space. The dispersion is calculated assuming each image has a gaussian intensity profile, and solving for the dispersion from the intensity at the center and threshold radius of the image. The size of the COSMOS spot results in a larger value of sigma than would correspond to plate seeing.

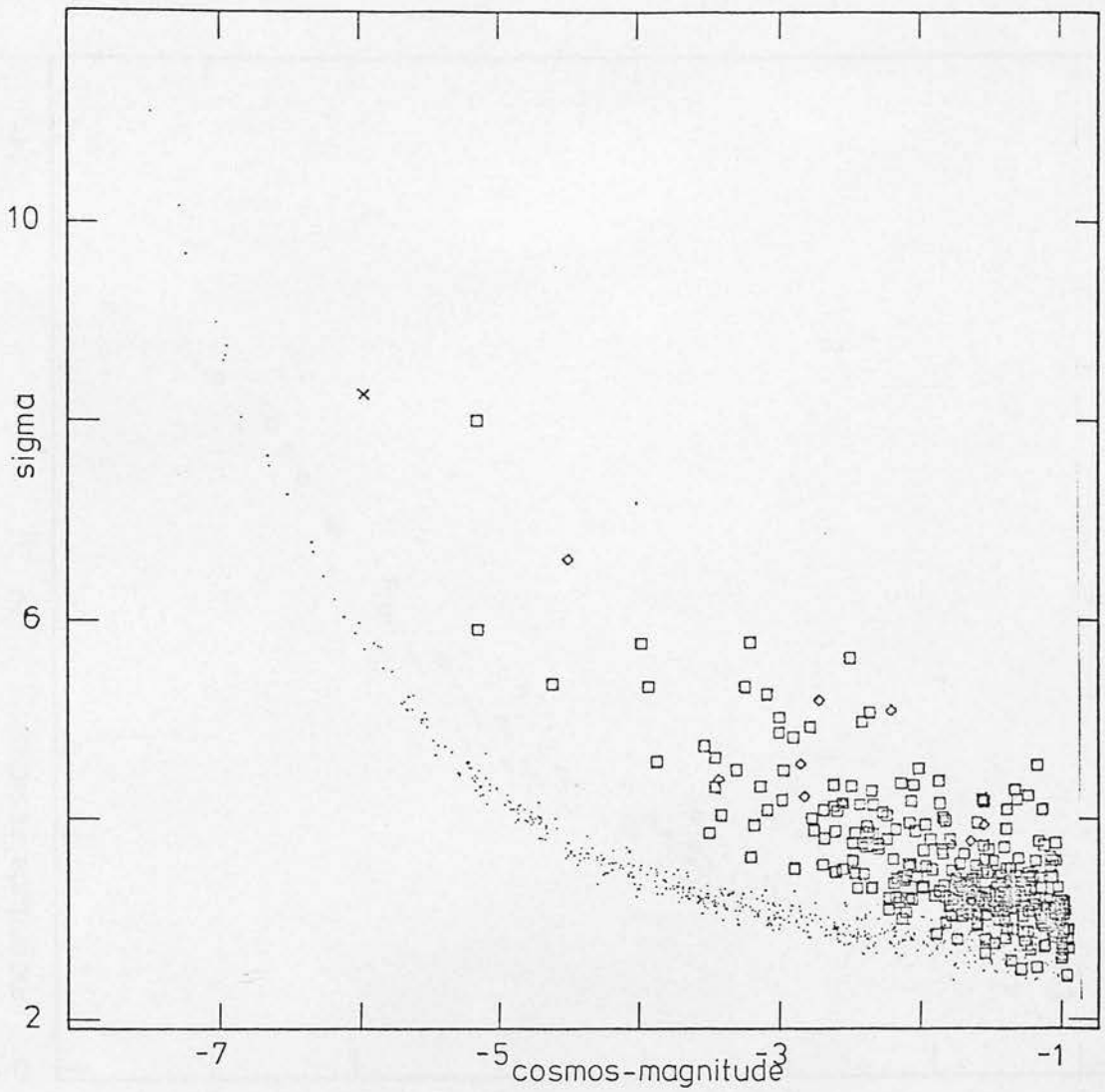


Fig. A1.5b

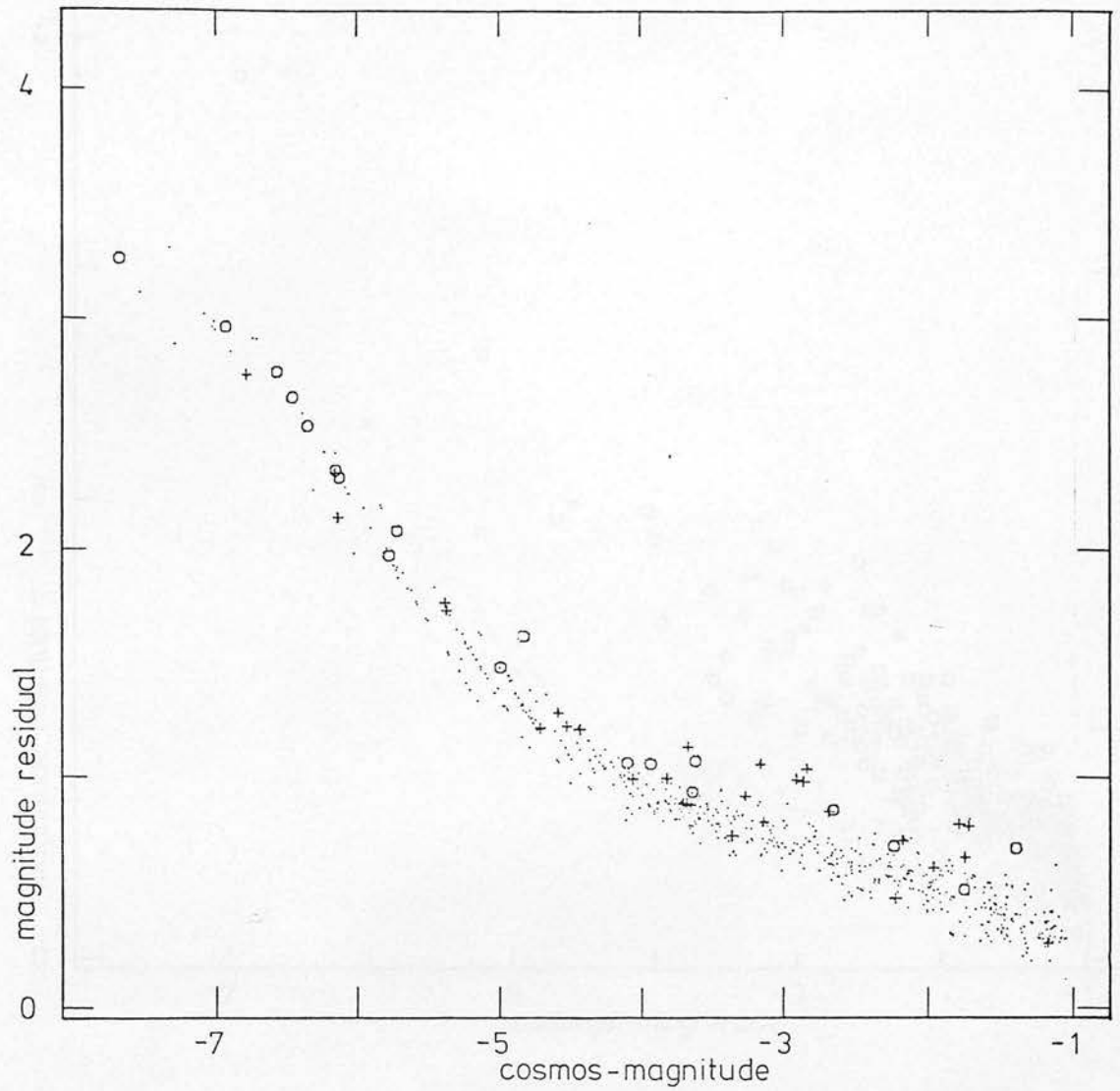


Fig. A1.6 The distribution of 1100 images from plate V3475 in cosmos-magnitude vs. magnitude residual from gaussian approximation space. The magnitude residual is calculated assuming a representative gaussian image profile; an integrated intensity is calculated for each image from the central intensity and threshold radius. The magnitude residual is that between the observed image magnitude and calculated gaussian magnitude. The parameter is similar to the Sigma parameter illustrated in Fig. A1.5, but the profile information is used in a somewhat different fashion.

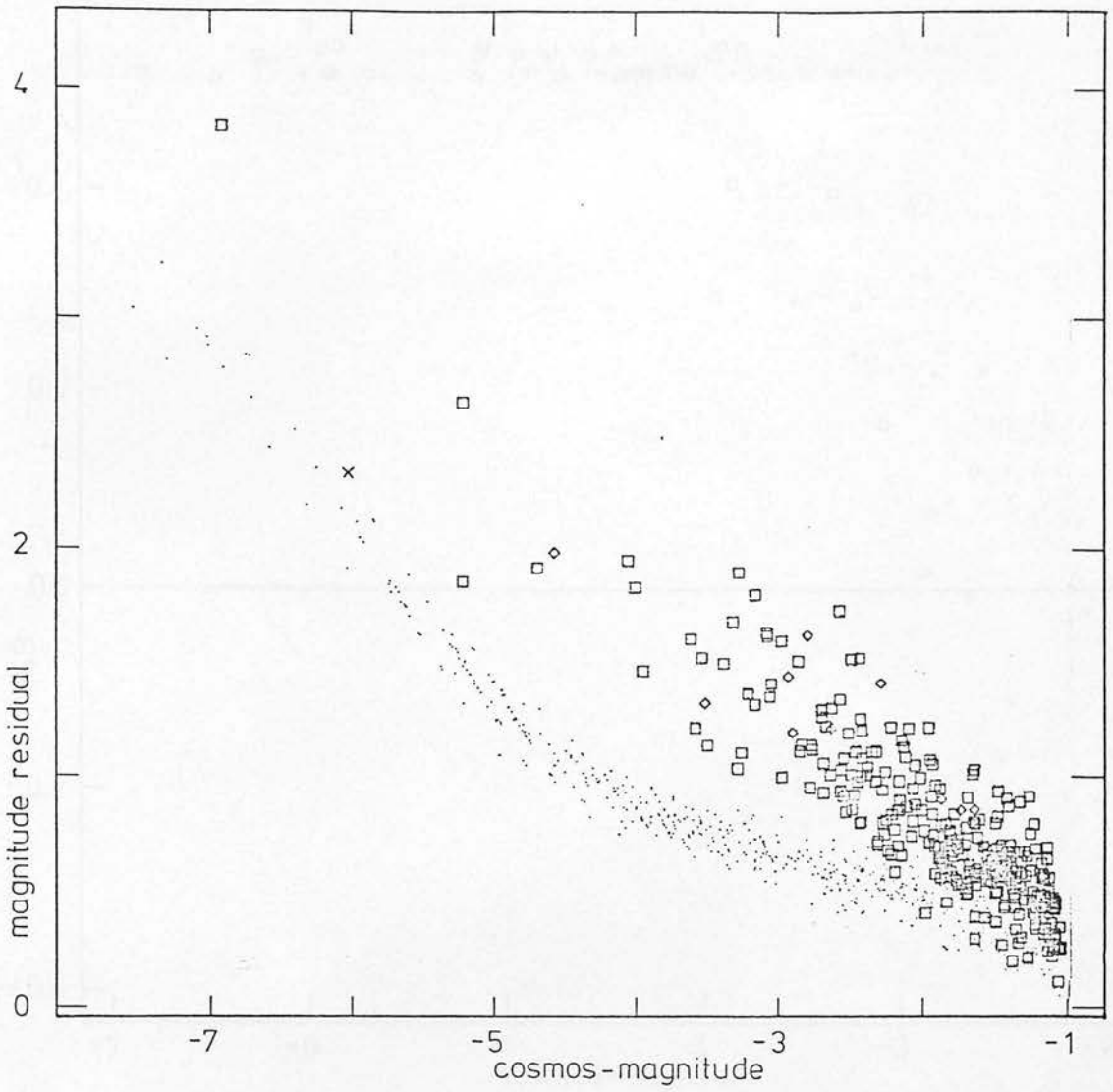


Fig. A1.6b

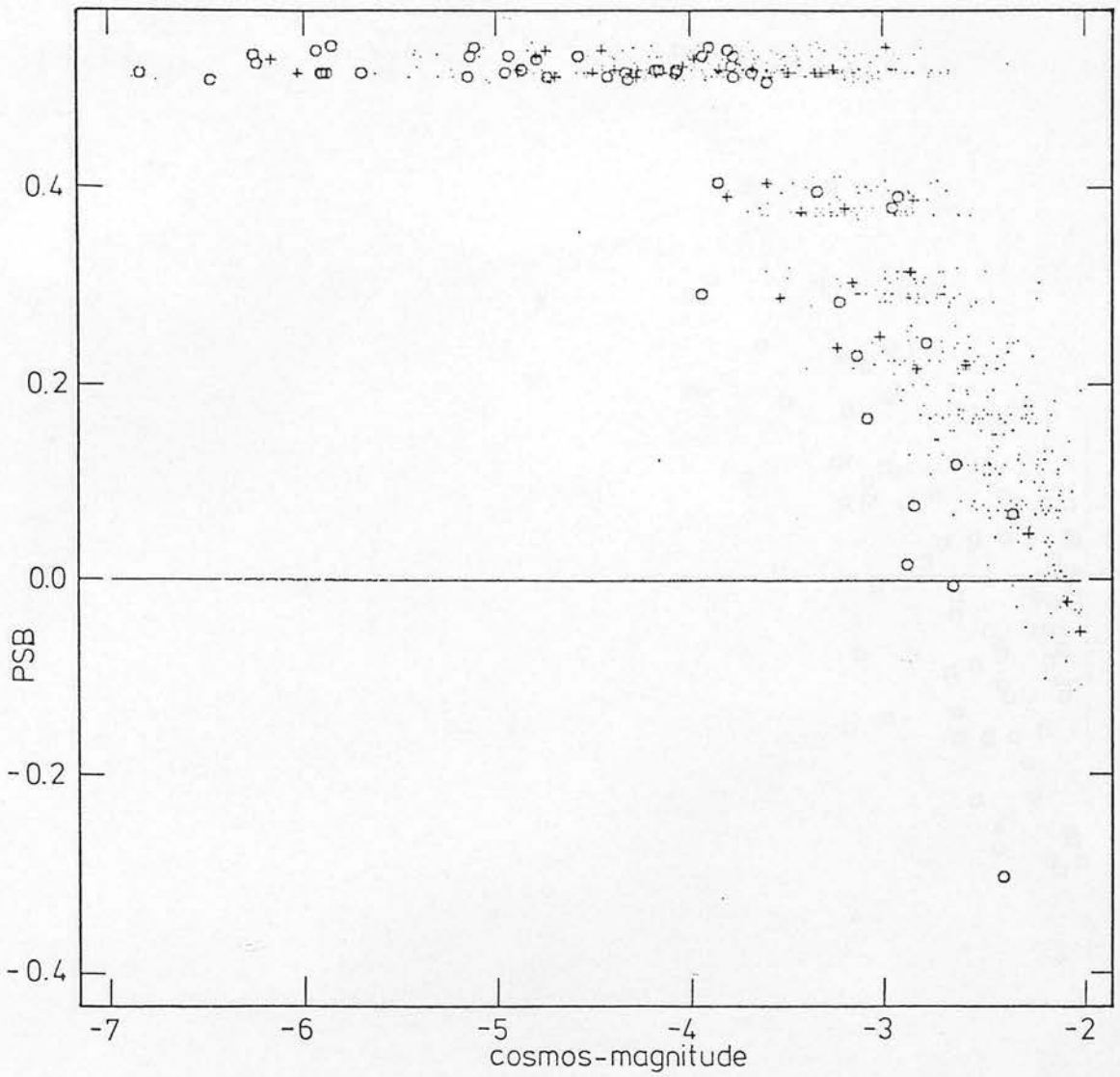


Fig. A1.7 The distribution of 750 images from plate J3001 in cosmos-magnitude vs. peak surface brightness space. The effects of saturation are far more evident compared to the corresponding figure for plate V3475 - Fig. A1.1.

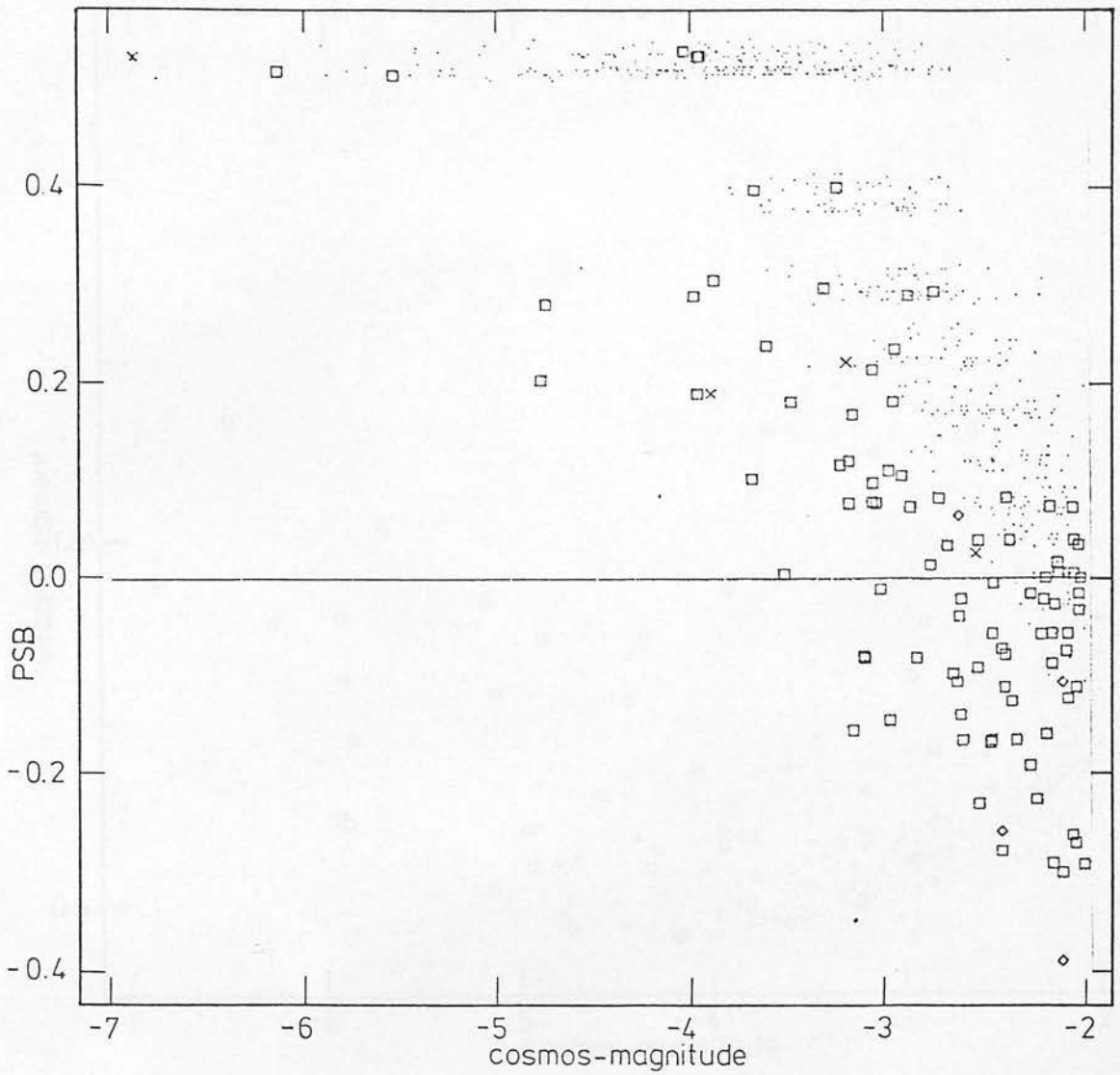


Fig. A1.7b

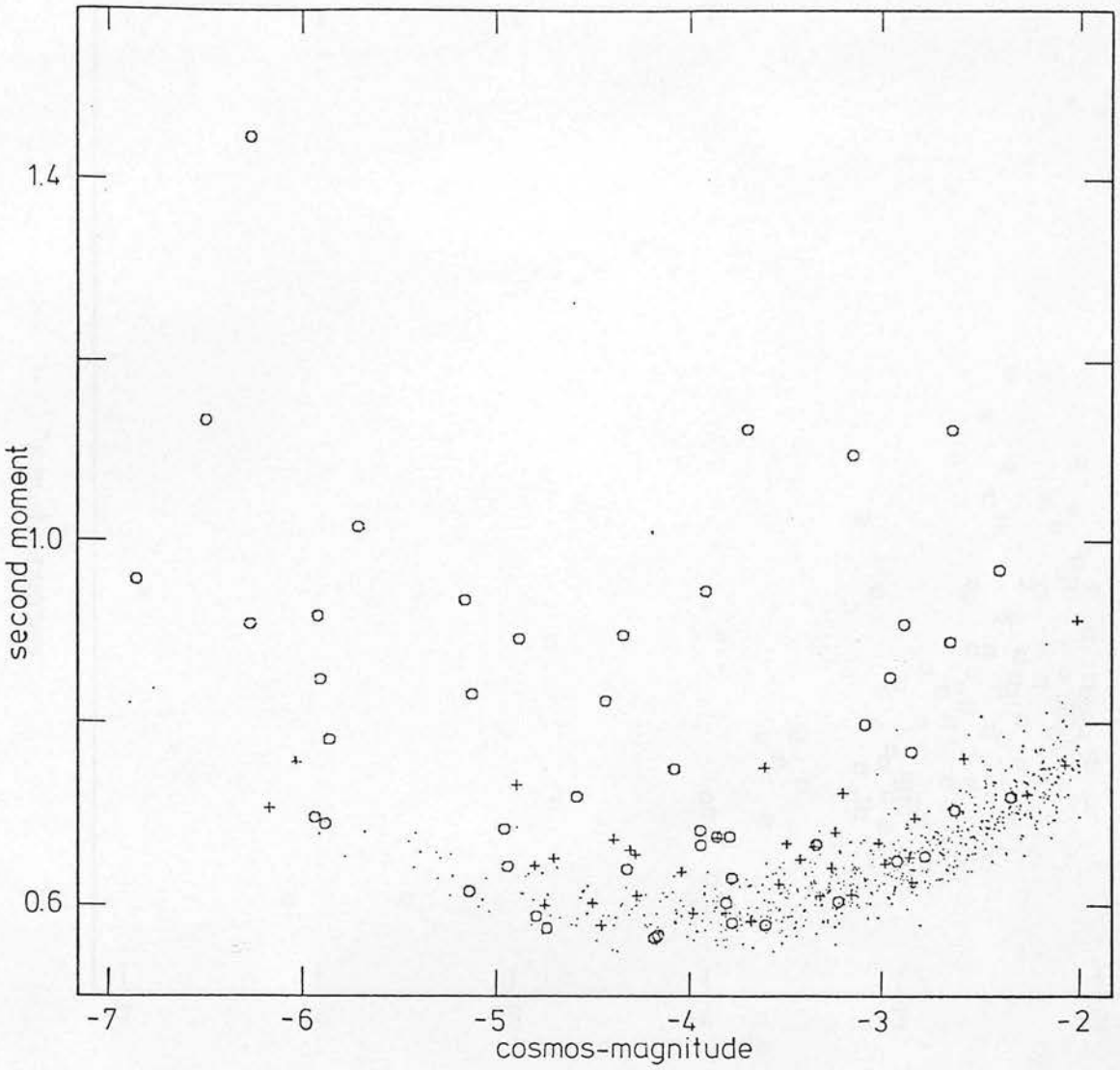


Fig. A1.8 The distribution of 750 images from plate J3001 in cosmos-magnitude vs. intensity weighted second moment space. The turn-up characteristic of stellar images in this space at bright magnitudes is due to the occurrence of diffraction spikes.

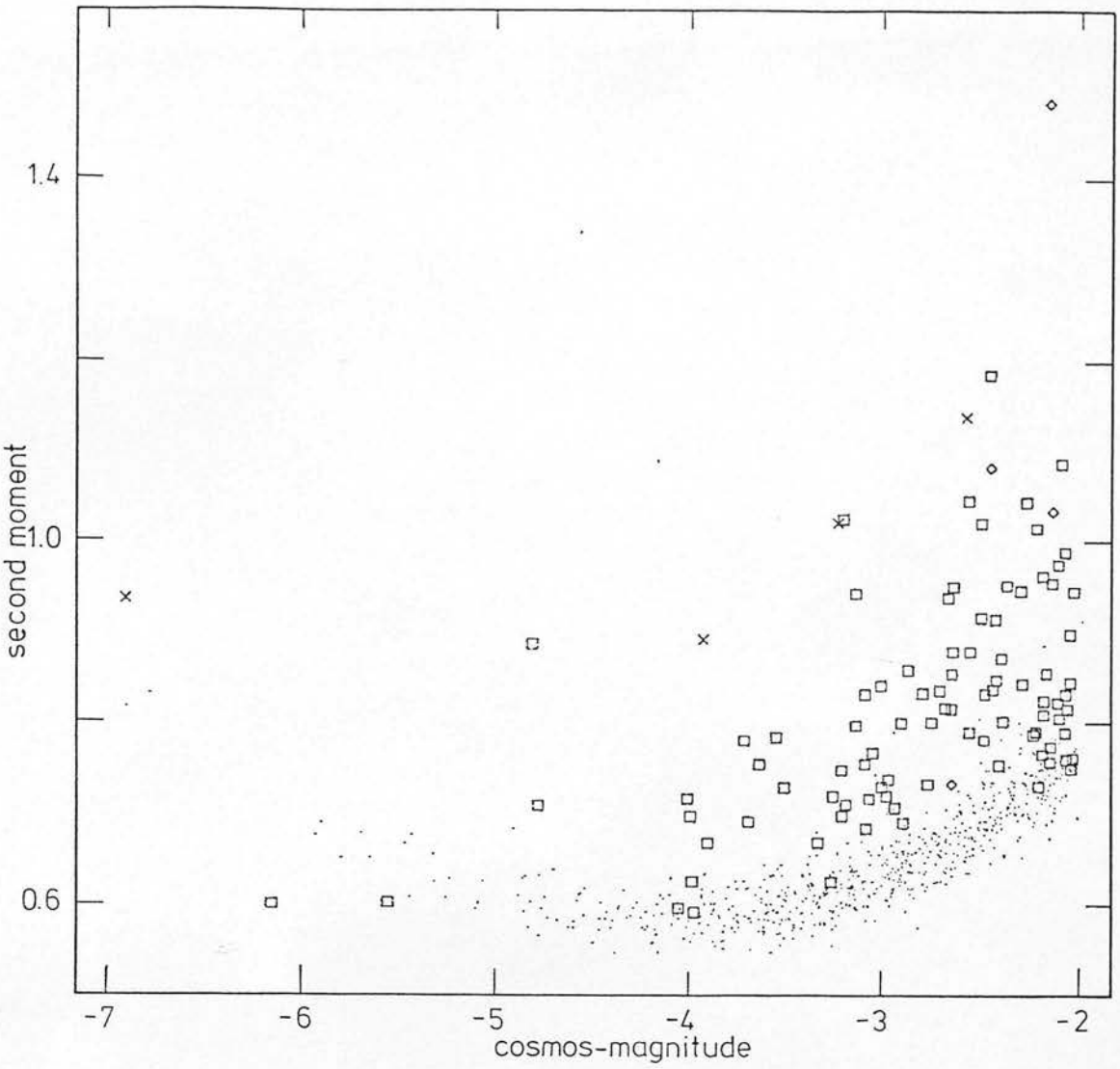


Fig. A1.8b